

THE SAGE ENCYCLOPEDIA OF SOCIAL SCIENCE RESEARCH METHODS

社会科学研究方法百科全书

【第三卷】

[美]米歇尔·刘易斯-伯克（Michael S. Lewis-Beck）
[美]艾伦·布里曼（Alan Bryman）
[美]廖福挺（Tim Futing Liao）—主编

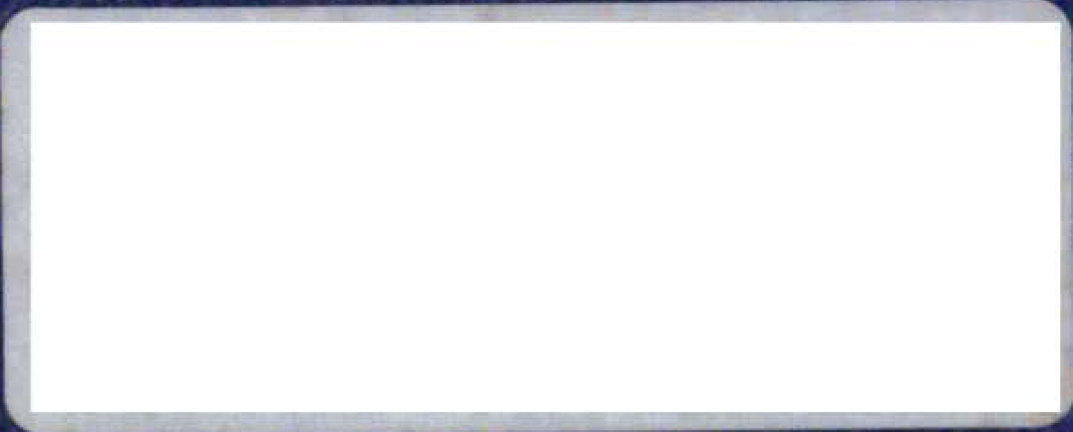
沈崇麟　赵　锋　高　勇　—主译

赵　锋　王　玥　马　妍　—本卷译者



重庆大学出版社

SHEHUI KEXUE YANJIU FANGFA BAIKE QUANSHU





更多服务


ISBN 978-7-5689-0785-9

9 787568 907859 >

总定价：698.00元（共三卷）



社会科学研究方法
百科全书

【第三卷】

[美]米歇尔·刘易斯-伯克（Michael S. Lewis-Beck）
[美]艾伦·布里曼（Alan Bryman）
[美]廖福挺（Tim Futing Liao） 主编

沈崇麟 赵 锋 高 勇 主译

赵 锋 王 玥 马 妍 本卷译者



译者（按姓氏拼音排序）

高 勇（中国社会科学院） 李 洋（北京市社会科学院） 马 妍（中国社会科学院）
沈崇麟（中国社会科学院） 王 玥（首都师范大学） 杨 可（中国社会科学院）
叶鹏飞（中国劳动关系学院） 赵 锋（中国社会科学院） 郑晓娟（北京大学）

词条审定专家（按姓氏拼音排序）

郭金华（北京大学） 郭志刚（北京大学） 何江穗（中国政法大学）
柯惠新（中国传媒大学） 李 原（中国社会科学院） 渠敬东（北京大学）
沈崇麟（中国社会科学院） 席仲恩（重庆邮电大学） 夏传玲（中国社会科学院）

重庆大学出版社

图书在版编目(CIP)数据

社会科学研究方法百科全书. 第三卷 / (美) 米歇尔·刘易斯-伯克, (美) 艾伦·布里曼, (美) 廖福挺主编; 沈崇麟, 赵锋, 高勇主译. --重庆 : 重庆大学出版社, 2017.8

(万卷方法)

书名原文: The SAGE Encyclopedia of Social Science Research Methods

ISBN 978-7-5689-0785-9

Ⅰ. ①社… Ⅱ. ①米… ②艾… ③廖… ④沈… ⑤赵… ⑥高… Ⅲ. ①社会科学—研究方法 Ⅳ. ①C3

中国版本图书馆 CIP 数据核字(2017)第199799号

社会科学研究方法百科全书

(第三卷)

[美]米歇尔·刘易斯-伯克 [美]艾伦·布里曼 [美]廖福挺 主编

沈崇麟 赵 锋 高 勇 主译

策划编辑:雷少波 柏子康 林佳木

责任编辑:雷少波 柏子康 屈腾龙 张慧梓 姜 凤 版式设计:雷少波

责任校对:邹 忌 责任印制:张 策

*

重庆大学出版社出版发行

出版人:易树平

社址:重庆市沙坪坝区大学城西路21号

邮编:401331

电话:(023)88617190 88617185(中小学)

传真:(023)88617186 88617166

网址:http://www.cqup.com.cn

邮箱:fxk@cqup.com.cn (营销中心)

全国新华书店经销

重庆升光电力印务有限公司印刷

*

开本:787mm×1092mm 1/16 印张:35.5 字数:816千

2017年8月第1版 2017年8月第1次印刷

ISBN 978-7-5689-0785-9 总定价:698.00元(共三卷)

本卷目录

Q

R

S

T

U

V

W

X

Y

Z

总目录

（按汉语拼音排序）

C

D

E

F

G

H

J

K

L

M

N

O

P

Q

R

S

T

V

W

X

Y

Z

读者指南

因果建模

话语/谈话分析

计量经济学

认识论

民族志

评估

应用定性研究
应用性研究
评估研究
实验
启迪调查法
影响评估
定性评估
随机对照试验

事件史分析

截删和删节
事件史分析
风险率
存活率分析
转移率

实验设计

实验
实验者期望效应
外在效度
田野实验法
霍桑效应
内在效度
实验室实验
米尔格拉姆实验
准实验

因子分析

聚类分析
共同性分析
验证性因子分析
对应分析
特征值
探索性因子分析
因子分析
斜交旋转
主成分分析
旋转因子
旋转
方差极大旋转

女权主义方法论

女权主义民族志
女权主义研究
社会性别议题
认识论立场

广义线性模型

一般线性模型
广义线性模型
连接函数
Logistic 回归
Logit
Logit 模型
泊松回归
Probit 分析

历史/比较

比较法
比较研究
文献类型
主位/客位之分
历史方法
口述史

定性研究中的访谈

传记叙事解释法(BNIM)
从属访谈
知情人访谈
定性研究中的访谈
叙事访谈
半结构化访谈
非结构访谈

多层次分析

情境效应
相依观察值
固定效应模型
混合效应模型
多层次分析
非参数随机效应模型
随机系数模型
随机效应模型

多元回归

调整的 R 方
最佳线性无偏估计量
Beta(β)
广义最小二乘
异方差性
交互效应
误设
多重共线性
多元回归分析
非加性
R 方
回归
回归诊断
设定
估计标准误

定性资料分析

分析式归纳
CAQDAS
持续比较法
扎根理论
原生编码
备忘录,撰写备忘录
不符案例
定性内容分析

定性研究的抽样

目的抽样
定性研究的抽样
滚雪球抽样
理论抽样

调查抽样

多阶抽样
配额样本
随机抽样
代表性样本
抽样
抽样误差
分层样本
系统抽样

度量

态度测量
双极量表
维度
双重度量
哥特曼度量
指数
李克特量表
多维度量(MDS)
最优度量
量表
度量
语义差异量表
瑟斯通度量

显著性检验

Alpha(α),检验的显著性水平
置信区间
显著性水平
单尾检验
检验的统计效力

简单回归

调查设计

时间序列

Q

Q 方法论(Q Methodology)

Q 方法论为研究主观性科学提供了一个框架,它结合了数据收集(Q 分类技术)程序和分析(因子分析【Factor Analysis】)程序。Q 方法论由威廉·史蒂芬森提出,首先需要就某个主题(如环境行动主义)汇集一个观点总体【Universe】,再从中选出 30~50 个陈述作样本,例如,"无论何种原因,也没有人可以凌驾于法律之上""我有一点怀疑他们的动机"等。接着,请二三十个参与者对这些题项依据同意(+5)到不同意(-5)予以排序。Q 序列是相关的,因此,可以使用 PQ 法或 PCQ 软件包(可从相关网站获取)作因子分析,从而揭示在研议题的主观多样性。同陈述关联的因子分构成了解释因子的基础。

例如,在环境行动主义研究中,将上述陈述分发给 36 位英国公众,他们的 Q 排序揭示出 7 个不同的叙述,例如,遵守法律、自由人文主义、激进的行动主义等(Capdevila & Stainton Rogers, 2000)。分离性障碍的单案例研究中,有多重人格的几个成员,就他们与小组中其他成员的关系,给出自己的 Q 排序。Q 排序的因子分揭示了人格的组织形态(Smith, 2000, pp. 336-338)。

如史蒂芬森的《心理学记录》(*Psychological Record*,1986—1988)在 5 篇系列文章中总结的那样,Q 方法论在概念和数学方面与量子力学相似。在人类努力的诸领域中,如文学、政治、决策、心理治疗、报纸阅读,主观行为都是概率的【Probabilistic】、不确定的、可传递的,但是当运用 Q 排序时,就有了明确的形式,其中的意义和测量都与 Q 分类者的判断行为密不可分。因子数目不是预先给定的,而是存在于互补的关系中。

Q 方法论产自行为主义【Behaviorism】领域,而后被后现代主义者、社会建构主义者【Constructionist】、女性主义者、话语【Discourse】和叙事分析者、认知科学家、心理分析学家、地理学家,以及定量和定性研究者广泛采用。它也被政策分析者采用,因为它可以揭示利益相关者(stakeholder)的视角(Brown, Durning, & Selden, 1999),还被那些关心新兴的民主认同的人所采用(Dryzek & Holmes, 2002)。在主观性科学研究国际协会的《操作性主观性》(*Operant Subjectivity*)杂志、《人的主观性》(*Human Subjectivity*)杂志,以及韩国主观性科学研究协会的韩语杂志《Q 方法论和理论》(*Q-Methodology and Theory*)上不断出现这方面的文章。

——Steven R. Brown

(王玥译　赵锋校)

参考文献

Brown, S. R., Durning, D. W., & Selden, S. C. (1999). Q methodology. In G. J. Miller & M. L.

Whicker (Eds.), *Handbook of research methods in public administration* (pp. 599-637). New York: Dekker.

Capdevila, R., & Stainton Rogers, R. (2000). If you go down to the woods today ...: Narratives of Newbury. In H. Addams & J. Proops (Eds.), *Social discourse and environmental policy: An application of Q methodology* (pp. 152-173). Cheltenham, UK: Elgar.

Dryzek, J. S., & Holmes, L. T. (2002). *Post-communist democratization*. Cambridge, UK: Cambridge University Press.

Smith, N. W. (2000). *Current systems in psychology*. Belmont, CA: Wadsworth.

Q 排序(Q Sort)

参见 Q 方法论【Q Methodology】。

二次方程(Quadratic Equation)

二次方程是用以求解未知项为乘方变量的初级代数方法。它是一种获得"根"(未知项与 x 轴的交点)的有用方法。

更确切地说,一个二次方程是一个二阶代数方程。它可能包含一个或多个变量,而单变量 x 的标准形式如下:

$$ax^2 + bx + c = 0 \quad (1)$$

式中,a,b 和 c 是方程的常数(也称为系数)。当系数为 0 时,方程可以简化;但当 $a \neq 0$ 时,二次方程仍然保持同样的形式,例如:

$$x^2 - 3x + 2 = 0 \quad \text{和} \quad 2x^2 - x - 3 = 0 \quad (2)$$

是典型的二次方程。一个二次方程的解通过解如下二次方求得:

$$x^2 + \frac{b}{a}x = -\frac{c}{a}$$

$$x + \frac{b^2}{2a} = -\frac{c}{a} + \frac{b^2}{4a^2} = \frac{b^2 - 4ac}{4a^2}$$

$$x + \frac{b}{2a} = \frac{\pm\sqrt{b^2 - 4ac}}{2a}$$

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad (3)$$

方程(3)被称为二次方程式,有助于求解未知项的方程根。

回到方程(2),二次方程式分析如下:

$$2x^2 - x - 3 = 0 \rightarrow a = 2, b = -1, c = -3$$

有解:

$$x = \frac{-(-1) \pm \sqrt{(-1)^2 - [4 \times 2 \times (-3)]}}{2 \times 2}$$

$$= \frac{1 \pm 5}{4} = \frac{3}{2} \text{或} -1$$

其中,算式 b^2-4ac 称作判别式,通常以希腊字母 δ 表示。二次方程的判别式决定解(也称为根)的性质,例如,如果

$$b^2 - 4ac > 0$$

二次方程有两个不相等的实根:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

反之,若

$$b^2 - 4ac = 0$$

方程仅有一个实根

$$x = -\frac{b}{2a}$$

若

$$b^2 - 4ac < 0$$

那么 x 没有实根,方程的解为复数根。

二次方程的根可通过对应二次曲线函数图来呈现。试看,

$$y = ax^2 + bx + c$$

式中,a,b 和 c 为常数,且 $a \neq 0$。x 对 y 的二次函数图称作抛物线,像碗形一样对称于一垂线。原二次方程($ax^2 + bx + c = 0$)的根是抛物线与 x 轴的交点处的 x 值。一个二次方程可能有一个(原二次方程有两个相等的根)、两个(原二次方程有两个不相等的实根)或没有与 x 轴的截距(原二次方程没有实根),如前文所述,不同的结果最终取决于二次方程的判别式。

二次方程更一般的形式如下:

$$ax^2 + by^2 + cxy + dx + ey + f = 0 \quad (4)$$

式中,a, b, c, d, e 和 f 是常数,且 a, b 和 c 中至少有一个不等于 0,例如:

$$4x^2 + y^2 - 4xy - 2x + y + 1 = 0$$

是 x 和 y 的一个一般二次方程。该二次方程的图称作圆锥曲线。它可以是一条双曲线、一条抛物线、一个椭圆形、两条交线、一个点、两条平行线或两条重合线。下述方法可用于确定一个一般的二次方程将产生哪一种类型的圆锥曲线(Zwillinger, 1995):

$$\Delta = \begin{vmatrix} a & \frac{c}{2} & \frac{d}{2} \\ \frac{c}{2} & b & \frac{e}{2} \\ \frac{d}{2} & \frac{e}{2} & f \end{vmatrix} \qquad J = \begin{vmatrix} a & \frac{c}{2} \\ \frac{c}{2} & b \end{vmatrix}$$

$$I = a + b \quad K = \begin{vmatrix} a & \frac{d}{2} \\ \frac{d}{2} & f \end{vmatrix} + \begin{vmatrix} b & \frac{c}{2} \\ \frac{c}{2} & f \end{vmatrix}$$

表 1 所示各类别表示了各参数值所对应的几何形式。

表 1

Δ	J	Δ/I	K	圆锥图形式
≠0	<0			双曲线
≠0	0			抛物线
≠0	>0	<0		椭圆线
0	<0			两条交线
0	>0			点
0	0		<0	两条平行线
0	0		0	两条重合线

通过增加额外的变量,二次方程可以变得更具一般性,但是,所形成的多维图形可能产生相当复杂的表面。

——Zhen Li

(马妍译　赵锋校)

参考文献

Zwillinger, D. (1995). *CRC standard mathematical tables and formulae* (30th ed.). Boca Raton, FL: CRC Press.

定性内容分析(Qualitative Content Analysis)

定性内容分析可用于检验文本的意义内容,当这些内容不适于定量技术时。这类定量技术测量了大量文本中的频次和规律模式,但是,正如西格弗里德·克拉劳尔(Kracauer, 1952-1953)在一篇重要论文中所论证,某篇文本的核心意义可能不适于量化。或者,文本的某些特征不能被轻易地计数,即便可以量化,也不能告诉我们诸特征在文本中或文本间如何发挥作用。反量的文本特征和功能包括:讽刺、两可、隐喻;交流语和陈述模式;民俗主题、美学符号、文学规范;修辞、风格,如共鸣、隐喻及其他言谈特征;以及可能隐含于信息之中,并使得某些概念或观念在意义上显得自然或绝对的立场、假设和价值。

举个例子,某些前提和价值预设会联合起来,改变一个特定的范畴,使其附属于其他范畴的原型和概念。于是,当分析一个原型或概念的构建、操作时,既需要形成言之有据的理论论证,也需要结合多种方法程序(Pickering, 2001)。

当统计描述达到其限度时,定性内容分析就登场了,其目的在于揭示重要的文本特征的意义——潜在或隐藏于表面内容之下,或揭示直接的、表面的内容的深层含义。即使是一个单独的词汇选择,例如,一份报纸上的一个标题或一份政府报告中的富有意涵的术语,都可能通过与文本的其他内容发生一系列相互关联,通篇发生作用。同时,一份特定的文本可能引出一条发展和反应的意义链,而在词语频次的定量模式中,则不能发现该意义链。对于克拉劳尔来说,文本的此类特质造就了它的暗指和隐义的结构。

研究媒介内容和文化文本的两种最重要的定性方法是符号学【Semiotics】和批判话语分析【Critical Discourse Analysis】(Deacon, Pickering, Golding, & Murdock, 1999, Chaps.7, 8)。两种方法均提供了各种可以打开一份档案或人造物品中的暗指和隐义结构的概念工具,从而显出在生产和传播它们的文化过程中,它们所具有的广泛意涵。两种方法还能显出,通过不同的意义和价值流,某些特定的意义和表象怎样被反复强调或交织在一起。此类形式的定性内容分析并不把文化文本当作孤立的实体,而是把它们当作意义生产过程和意义阐释过程的互动场所。正因如此,文化文本虽然具有确定生产印记,但是其意义却不固定。它们给出了阐释的线索,并允许多种解读。定性分析最适于追踪上述视角(文化文本从其开始)和导向(引向它们被建构和重复建构的方向),以及由此产生的意义——它们一贯具有的、暂时性的意义内容,即便它们总是看上去要求一种固定的、唯一的、合法的解释。

——Michael J. Pickering

(王玥译 赵锋校)

* 也可参见内容分析【Content Analysis】。

参考文献

Deacon, D., Pickering, M., Golding, P., & Murdock, G. (1999). *Researching communications*. London: Arnold.

Kracauer, S. (1952-1953). The challenge of qualitative content analysis. *Public Opinion Quarterly*, *16*, 631-642.

Pickering, M. (2001). *Stereotyping: The politics of representation*. Basingstoke, UK: Palgrave.

定性资料管理(Qualitative Data Management)

在一个成功的定性研究中,资料管理是其中一个有挑战性的、不可或缺的、关键的组成部分。好的资料管理被认为是促进项目连贯有序的必备条件(Huberman & Miles, 1994)。管理好资料“有益于解释,就好像好的配器法有助于好的舞曲”(Meadows & Dodendorf, 1999, p. 196)。从一开始,研究者就要有一个计划,去分类、总结、分析和储存项目资料,还需要把这些工作的重复进行也包括在计划中。定性研究的往复前进意味着资料管理和资料分析是互不可缺的。

定性研究中的资料管理始于项目的概念化。尽管依据既有传统,构成资料的内容会有所不同,但是,做定性研究,并做好它,则意味着研究者的反思是项目资料的组成部分。根据反思和现有文献(和潜在理论)的考察,围绕某个兴趣主题,形成研究问题的系统表述是项目的核心。一个全面的文献回顾是定性研究的第一步(Morse & Richards, 2002)。研究问题指明将要收集的资料:没有好的资料,就不可能有好的研究。在研究设计阶段,关于抽样、调查对象征召、资料收集技术以及分析过程所做的决策是资料管理中的一个个步骤,其中每一步都需要记录下来,归置入项目资料。

资料的来源各种各样,形式也相当丰富,包括文档、图片、新闻片段、文字转录,以及研究初期的田野笔记和反思性日志。这些资料需要尽早且经常地存档、转成可溯源的文本,或其他可以被分析和使用的形式。研究者必须记得备份项目中使用的所有资料。编码、类别或主题的描述,以及反映了资料从描述到分析前进过程和最终的解释及/或理论构建过程的备忘录,也都是一个项目中所有资料的组成部分。

定性分析软件是这个过程的必备组成,可以以下述方式助益于研究项目的管理,即容许研究者把时间花在资料的分析和解释上,而不是用在寻找重要的引述或寻找带有重要编码却难以查找的文件上。资料可以放在文件柜的抽屉中手工管理,但是示意图、信息图表、模型和其他表格可能难以如此管理,否则会妨碍研究团队与资料的紧密联系。分析软件还可协助对已编码的资料提出问题、发展理论,或是有助于解释过程。

最后,管理还包括坚守伦理标准,其中包括资料的匿名化,确保只有研究团队成员可以接触资料,还要确保在项目全过程中各种形式的资料都被安全地保存。在研究项目起始阶段,仔细地规划资料管理可以对定性研究的严谨和成功作出重要的贡献。

——Lynn M. Meadows

(王玥译　赵锋校)

*也可参见数据管理【Data Management】。

参考文献

Huberman, M. A., & Miles, M. B. (1994). Data management & analysis methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 428-444). Thousand Oaks, CA: Sage.

Meadows, L. M., & Dodendorf, D. M. (1999). Data management and interpretation: Using computers to assist. In B. F. Crabtree & W. L. Miller (Eds.), *Doing qualitative research* (pp. 195-218). Thousand Oaks, CA: Sage.

Morse, J. M., & Richards, L. (2002). *Read me first for a user's guide to qualitative methods*. Thousand Oaks, CA: Sage.

定性评估(Qualitative Evaluation)

判断一个项目的有效性时,评估者常使用开放式访谈、观察、文献回顾和个案研究方法。这些方法构成了定性评估。

定性评估强调用项目参与者自己的语言来报告他们的经验。评估者通常收集统计数据,但是理解实际构造数字的人的故事和看法也很重要。定性评估不仅要评估目标的达成,还要了解那些意想不到的影响和连锁反应,还要阐明预期结果中难以量化的部分,如对参与者来说实现目标意味着什么。

定性方法尤其适用于成果因人而异的项目。这些项目提供的服务能满足个别参与者的不同需要。在此类项目中,对于不同服务对象,成果可以有性质不同的内容。定性评估则可通过单个的个案研究,记录各个结果的变化。

过程评估通常使用定性方法来研究一个项目如何产生成果。过程评估旨在通过调查以下几种问题来解释和理解项目操作的内在动力:人们经历了什么事情让项目变成现在这样?服务对象如何被带进项目,以及一旦成为参与者他们是如何经历这个过程的?人们所做的与他们想要完成或实际完成的有何关联?从参与者和项目工作者的视角看项目的优势和劣势是什么?

定性评估非常适合研究项目的实施过程,因为:(a)描述过程要求细致地描述人们如何参与其中;(b)参与者的参与过程的经验通常是不同的,所以需要抓住他们不同的视角;(c)过程是流动、动态的,不可用某一时点的简单的等级标准来概括。过程评估要考察正式的活动和非正式或非计划的互动。

定性方法也可用在执行评估中,以记录一个项目是如何发展,以及项目如何又为什么会偏离初始计划和期望。其他评估关心的是哪一种定性方法特别适合,包括比较不同项目、记录历时发展、调查系统变迁,以及个人化和人性化的评估(Patton, 2002)。

参与式评估常常严重依赖于定性方法,因为它们比统计数字更容易理解。在参与式评估中,评估者与项目工作人员和/或参与者一起来评估项目的有效性。开放式的访谈是参与式评估的一种常用方法,因为参与者可以在评估者的帮助和指导下学习收集和分析资料。将参与者吸收入合作之中可以增加评估的实用性(Patton, 1997)。

定性方法既可以单独使用,也可以通过混合研究方法设计(mixed-methods design),结合定量技术一起使用。尽管有过争议,但是定性方法在评估研究中不仅已经被广泛应用,其应用价值也不断提升。

——Michael Quinn Patton

(王玥译　赵锋校)

*也可参见评估研究【Evaluation Research】。

参考文献

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: Sage.

Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage.

定性元分析(Qualitative Meta-Analysis)

定性元分析,也指定性元综合(qualitative metasynthesis)(Sandelowski, Docherty, & Emden, 1997)、定性元资料分析(qualitative meta-data-analysis)(Paterson, Thorne, Canam, & Jillings, 2001)和元民族志【Meta-Ethnography】(Noblit & Hare, 1988),它是一种特别的综合类型,即来自某个研究领域的所有定性研究的研究发现被正式地组合在一起。类似于定量元分析,定性元分析既是一项分析过程,也是一项阐释成果,其目的在于系统地、全面地、清楚地查明某一研究领域的知识状况。作为一种研究方法,部分由于其早期的发展,定性元分析被描绘成一种内含于或等价于其他类型定性综合的形式,例如,扎根理论【Grounded Theory】、元研究和定性资料的二次分析【Secondary Analysis of Qualitative Data】。

开展定性研究的兴趣之所以逐渐增长,其主要因素在于健康、社会科学、实践学科中大量定性研究的涌现。与之相应,得自这些研究中的重要发现并未充分地用于实践和政策讨论。此外,研究者也迫切地想把描述的证据和以证据为基础的实践完全当作实验研究的结果。由于强调通则的知识甚于独特的知识,强调形式的一般甚于有限的个案,定性元分析旨在提升定性研究的关联和效用。因为,相较于任意定性研究所遵循的典型样本规则,定性元分析置细致的、个案取向的现象研究于更大且更富于变化的样本中。然而,定性元分析还要保存个别研究中具体的经验。尤其在实践和政策讨论中,一般通则在用于个别情形时经常失效,而细节的知识则可作为一般通则的有效补充。

正是定性研究强调细节才造成整合研究结果的巨大困难。定性研究方法的多种多样、定性研究议程的品类繁多也使定性研究发现的综合变得复杂。实施定性元分析的其他挑战包括:界定研究的实质限度;获取相关研究;确定方法的诸特性是否,以及在多大程度上影响研究发现;判断一项研究的质量;确定一项研究是否属于定性研究;基于要报告的定性研究风格上的不同,发现其中的结果。此外,问题还包括:哪种分析的、解释的、呈现的技术最适于综合各个研究的发现;应当如何向各种受众汇报定性元分析的结果。

尽管存在诸多挑战,但整合的努力在增长。最显著的标志是:定性元分析研究的出版物不断增多;相关的研究基金持续增加;坎贝尔和科克伦研究综述数据库(Campbell and Cochrane databases of reviews of research)中纳入了定性研究。

——Margarete Sandelowski

(王玥译　赵锋校)

* 也可参见元分析【Meta-Analysis】。

参考文献

Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography: Synthesizing qualitative studies*. Newbury Park, CA: Sage.

Paterson, B. L., Thorne, S. E., Canam, C., & Jillings, C. (2001). *Meta-study of qualitative health research: A practical guide to meta-analysis and meta-synthesis*. Thousand Oaks, CA: Sage.

Sandelowski, M., Docherty, S., & Emden, C. (1997). Qualitative metasynthesis: Issues and techniques. *Research in Nursing & Health*, 20, 365-371.

定性研究(Qualitative Research)

定性研究是一个泛称,包含了一系列的实施质询(inquiry)的态度和策略,而这些质询旨在探究人类如何理解、体验、解释,以及生成社会世界(Mason, 1996)。尽管定性研究通常与定量研究并列,然而定性研究并非是一种明显有别于定量研究【Quantitative Research】的统一的调查研究的形式,而更像是来自科学、人文和实践学科的不同学者的家园,这些学者致力于不同的哲学和方法论立场,有时这些哲学和方法论立场可能是冲突的。

这些哲学立场包括阐释主义【Interpretivism】、诠释学【Hermeneutics】、社会建构论【Social Constructionism】,以及批判理论【Critical Theory】,但不限于此,因为这些哲学立场也有不同的界定。方法论的途径包括现象学【Phenomenology】、扎根理论【Grounded Theory】、民族志【Ethnography】、参与式调查、叙事分析【Narrative Analysis】和话语分析【Discourse Analysis】,也不限于上述几种,因为它们的界定也多种多样(Denzin & Lincoln, 2000a)。历史、文学批判、伦理学、文化和物质文化研究,这些学科通常被完全界定为定性研究。然而,如人口史,大部分是定量研究,这个领域的学者通常不认为他们自己是定性研究者,也不认为他们的工作是定性研究,甚至根本不认为是研究。社会科学领域的定性研究者转向艺术和人文学科,学习如何像艺术家和人文学者那样去阅读和完成文本;而人文学科的定性研究者转向社会科学,学习如何像社会科学家那样理论化以及如何开展实地调查研究。简言之,定性研究是一个在其差异性和同质性上都很模糊的术语。简单地通过与定量研究作对比,是不能理解定性研究的,而实际上我们经常这样做,这样会误解和曲解定性研究。相应地,试图减少定性研究的多样性和复杂性,这种做法同样也会曲解定性研究。

定性研究与定性资料和技术

应当将定性研究同那些仅仅包含定性资料的其他研究和其他工作区别开,例如,描述性的抽样调查【Sampling Survey】和新闻报道。还应当将定性研究同那些通常作为定性研究技术的资料收集和分析技术区别开,例如,焦点小组【Focus Group】和内容分析【Content Analysis】。虽然焦点小组和内容分析都源自定量研究,且依然在定量研究中有广泛的使用。然而,在这些技术中,定性研究的基本观点和策略并不是其内在构成。此外,所有关于人类主体的行为科学和社会科学的定量研究都要依赖所用问卷【Questionaire】中的定性资料。因此,仅仅由于这一事实,即某个项目中使用了字词,或者使用字词多于数字,并不能使该项目成为一个定性研究,就像是仅凭如下事实,即某个项目中使用数字或生产数字,也不能使该项目成为定量研究。

界定定性研究的特征

一个称得上定性研究的研究,它明显地致力于创造丰富的、相对细节性的描述,特别是对于人的、社会的、语言的、物质的,以及其他由这些构成的实践和事件的阐释。定性研究通常包括辨明人们的视角,或是那些常被称作行动者的立场的东西,当然不限于此。尽管在哲学和方法论上,定性研究是一种相当多样化的实在,但是某些确定的必要特征使其显著不同于其他研究,包括它的个案(与变量相反)导向,对于文化和历史背

景的敏感,以及反身性【Reflexivity】。在它的许多面向下,定性研究是一种经验质询的方式,而这一方式常含有下列必要程序:某种形式的目的抽样【Purposive Sampling】,以获得信息丰富的个案;深度访谈【In-Depth Interview】和开放式访谈,长期参与/实地观察,以及/或者档案研究或实物研究;还有相关的资料分析和阐释技术——能超越所得资料【Data】和资料的表层意义。定性研究通常趋向于生产特殊性【Idiographic】(与普遍性【Nomothetic】相反)的知识,因为它强调的是对特别现象、事件或个案的富有洞察力的理解。

相较于定量研究的共有特征——将客观主义【Objectivity】原则贯彻于他们对研究对象的态度和收集与处理数据的态度之中,定性研究则追求受训的主观主义,即它拥抱偏见——即便要清楚地说明所有质询中固有的偏见。在所谓的定量研究的共有特征中,重点是对于研究条件的控制——使得那些能影响研究发现的效度的偏差和威胁最小化,以及在进入研究的调查实地之前,先致力于发展研究设计【Research Design】。相反,一般定性研究的重点是强调自然主义【Naturalism】,此即,在不操控任何条件的情况下,观察事件以及进入调查研究的实地之后,对研究设计的持续发展和改进。正如梅森(Mason,1996)深刻地指出,定性研究要求人们能够“有策略地思考和行动”,与此同时,能够自由地、富有创造性地“将智力的、哲学的、技术的、实践的(伦理的)考量结合起来”(p. 2)。定性研究的特征还包括坚持一组多样化的研究取向和研究策略,从而使得研究的程序和结果取得最大的效度【Validity】或可信度,以及使用有表现力的语言和更富有文学表现力的风格来发布研究发现。

多种方式,一种基调

尽管我们不赞成用简单的定义来囊括定性研究一词中丰富的共同特征,但是,它依然传递了一个“基调”(Sapir,1951,p.308)。这一“基调”崇尚个别的、多样的、日常的生活。这基调还希望在被分隔开的人之间建立起桥梁。

——Margarete Sandelowski

(王玥译　赵锋校)

参考文献

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2000a). *Handbook of qualitative research* (2nd ed.). Thousand Oaks, CA: Sage.

Denzin, N. K., & Lincoln, Y. S. (2000b). The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 1-28). Thousand Oaks, CA: Sage.

Mason, J. (1996). *Qualitative researching.* London: Sage.

Sapir, E. (1951). Culture, genuine and spurious. In D. G. Mandelbaum (Ed.), *Selected writings of Edward Sapir in language, culture and personality* (pp. 308-331). Berkeley: University of California Press.

定性变量(Qualitative Variable)

一个变量的值随其属性和特征而变化,即称它为定性变量。定性变量的测量尺度是一组无序的(unordered)或名义的(nominal)类别。例如,性别(女性、男性);党派身份(民主党、共和党、无党派);婚姻状况(单身、已婚、离婚、丧偶);种族(白

人、黑人、西班牙裔、亚裔和其他);就业状况(未就业、在业、退休);以及宗教归属(天主教、犹太教、新教、伊斯兰教、佛教、无宗教信仰),都是定性变量。

不同于定量变量,定性变量不是根据数量或大小,而是依照性质或属性来分类。换句话说,不假定任一类别大于或小于其他任一类别。例如,"男性"和"女性"仅仅是性别的标签。这两个标签不表示数量和大小方面的差异。类似地,"卫理公会""天主教""长老派""伊斯兰教"和"佛教"是识别某人宗教归属的标签。它们都不表示变量"宗教归属"的数量是大还是小。

在社会研究中,人们可以根据某种特征用定性变量来对数据进行分类,也可以判断哪些元素最相似和最不相似。因此,定性变量的测量层次是定类尺度。术语"定类(nominal)"意味着类别间没有数值差异。如果一个变量在一个样本【Sample】或总体【Population】的不同主体有数值上的变化,那么这个变量是一个定量变量,而非定性变量。定量变量【Quantitive Variable】依定距尺度测定,因为同一尺度的任一数据对间有数值差或间距。例如,一个被访者的年龄可以用定距尺度来测量。30 岁和 50 岁的间距是 20 岁。因此,可以通过比较两个人的年龄来判断一个人比另一个大多少或小多少。然而,对定性变量而言,此类比较既不可能也无意义。

我们很有必要了解位于定距和定类尺度间的第三种测量层次。如果一个变量的值在定序尺度上具有自然的排序,但在任一对类别间没有确定的间隔距离,就称它为一个定序变量。例如,可以用非常自由的、比较自由的、温和的、比较保守的和非常保守的排序来测定一个人的政治意识。这个有序类别的集合具有一个明确的顺序,但是类别之间的距离无法确定。比如,尽管我们知道一个认为自己"非常自由的"人比一个被归为"保守"的人更自由,但是我们无法获知那个人更加自由的具体数值。

定序变量和定性变量通常被合称为分类变量。然而,分析定序和定性变量的统计方法通常是不同的。数据分析者需要知道定距-定序-定类尺度分类法和定量-定性分类法,因为每一种统计方法都对应特定的数据类型。例如,总计一个变量的均值【Mean】和标准差【Standard Deviation】仅适用于定量数据,而不适用于分析定性数据。我们注意到,研究者常出于编码【Coding】的目的,把离散数字分配给一个定性变量的类别和标签(例如,白人=1,黑人=2,西班牙裔=3,亚裔=4,其他=5)。然而,赋予数据特定类别的数字不能像分析定量变量和定序变量一样去计算。关于分类【Categorical】变量统计方法更详尽的内容可以在艾伦·阿格雷斯蒂(Agresti, 2002)、斯科特·朗(Long, 1997)、丹尼尔·鲍尔斯和谢宇(Powers & Xie,2000)的书中找到。

——Cheng-Lung Wang

(马妍译　赵锋校)

参考文献

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley Interscience.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.

Powers, D., & Xie, Y. (2000). *Statistical methods for categorical data analysis*. San Diego, CA: Academic Press.

分位数(Quantile)

一个分位数是低于给定值的数据【Data】点的比例(或百分比)。例如,0.25(或25%)指25%的观察值位于该点之下,同时,75%的观察值高于这个值。分位数有许多应用,既可用于分位数-分位数图(*p-p*图),也可用于分位数回归【Regression】。

——Tim Futing Liao

(马妍译　赵锋校)

定量和定性研究论争(Quantitative and Qualitative Research, Debate About)

定量与定性研究的论争所考虑的问题是:两种研究策略是否应该被视作对立的认识论立场,或者它们是否可以更好地被定义成,仅指两种不同的资料收集和资料分析的技术簇。这个问题对于多策略研究【Multistrategy Research】的前景有着非常重大的意义。如果定量和定性研究被视作特别的且极不协调的认识论,多策略研究就难以想象。另一方面,如果它们被视作资料收集和分析的不同方法,结合定量和定性方法的研究(如多策略研究)就比较容易设想了。

这一论争的认识论说法有两个主要的形式(Bryman, 2001)。一个是嵌入方法(embedded methods)说。这一论点断言,任何研究方法都植根于认识论承诺(epistemological commitments)。因此,选择使用自填式问卷【Self-Administered Questionnaire】做研究,不只是用某种方式来寻求研究问题的答案;这个选择同时伴随着对与之相关的(如实证主义【Positivism】)认识论(及本体论)原则的一个承诺。当休斯(Hughes, 1990)写道:"每一个研究工具或程序都无可避免地嵌入在对特定的世界观以及对那个世界认识的承诺中"(p.11),他便是彻底地采用了嵌入方法说。

关于这一论争的另一种认识论说法就是范式说。借用库恩(Kuhn, 1970)的范式【Paradigm】概念,这一论点将定量和定性研究描绘成不同的范式。如此一来,认识论的假设、价值和方法都被认为与两种范式密不可分。另外,因为相信范式不能调和,定量和定性研究即不可调和。这种认识论说法与嵌入方法说一致认为,研究方法联系于认识论立场,但是这种认识论说法则进一步把定量和定性研究看成是不可比的。

尽管关于定量和定性研究的论争的认识论说法还有拥护者——他们反对两种研究策略可以在更深层次结合在一起的观念(如Sale, Lohfeld, & Brazil, 2002),但是这种立场逐渐让位给论争的技术说。技术说的作者承认,定量和定性研究与认识论和本体论假设有关联,但并不将它们的关联视为必然的。换言之,研究方法可以服务于不同的主人,因此"出自一种研究策略的研究方法被认为可以塞入另一种研究策略中"(Bryman, 2001, p. 446)。

多数作者已不再坚持定量和定性研究的极端对立,逐渐接受争论的技术说,而这无疑促进了多策略研究的增长和接受。

——Alan Bryman

(王玥译　赵锋校)

参考文献

Bryman, A. (2001). *Social research methods*. Oxford,

UK: Oxford University Press.

Hughes, J. A. (1990). *The philosophy of social research* (2nd ed.). Harlow, UK: Longman.

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.

Sale, J. E. M., Lohfeld, L. H., & Brazil, K. (2002). Revisiting the quantitative-qualitative debate: Implications for mixed-methods research. *Quality and Quantity*, *36*, 43-53.

定量研究(Quantitative Research)

可以将研究分成两类:定量研究和定性研究。二者的区别在于,用什么方式收集数据,以及数据有多少观测值。抽样调查是定量研究的实例,而小规模的民族志个案研究是定性研究的实例。

一项定量方案的特点是,有一个研究者期望从中获得结论的总体,但又不可能收集整个总体的数据。就一项观测研究而言,必须选出一个适量的、统计随机样本【Random Sample】,并运用统计推论【Statistical Inference】的方法,获得相关总体的推论。就一项实验研究而言,要使用统计推论的方法,必须对进入实验组和对照组的受试者进行随机分配【Random Assignment】。

在定量研究的三个阶段都要运用统计方法。首先,在观测研究中,收集数据要运用统计抽样理论;然后,运用描述性统计分析来分析样本数据;最后,运用统计推论获得从样本数据到整个总体的一般推论。在实验研究中,首先,运用随机化方法分配受试者到实验组和对照组;然后,运用描述性统计分析去分析实验数据;最后,类似于分析观测数据,作出更大的总体的一般推论。

历史记录

在20世纪,抽取统计随机样本的理论和方法都有了长足的发展。英国的统计学家罗纳德·弗希尔是在收集数据时应用随机性的先驱,特别是在实验时。在实验研究中,有关实验进行的设计领域也有了较大发展。每种实验设计都有自己的计算方法,并且,在统计教科书和统计软件中,它们都更便于应用了。

20世纪后半叶出现了大规模的、国家级的抽样研究。由乔治·盖洛普及其合作者开展的研究是这类研究的著名实例,当然,还有许多同类的早期研究,Roper 系列研究就是其中之一。广为人知的学术研究包括:由最初成立于芝加哥大学的全国民意研究中心(National Opinion Research Center, NORC)和密歇根大学社会研究所的调查研究中心完成的系列研究。NORC 通过综合社会调查(General Social Survey)收集了多年的数据;密歇根的研究群体则在经济行为和选举投票的研究上有较长的历史传统。

应用

抽取一个回答者的全国样本是一项艰巨的任务,因为不存在一份可用的全国居民名录。当下,主要使用各种比较成熟的电话抽样方法,同时一系列专门的统计公式也被开发出来,用以分析此类数据。在所有介绍性统计学课本和多数统计软件包中的公式都以回答者的简单随机抽样【Simple Random Sampling】作为前提。但是,当样本是通过应用更复杂的方法获取时,例如用于大规模的、全国研究的样本,根据简单随机抽样而计算得到的标准误【Standard Errors】是不正确的。

例子

在运用抽样【Sampling】或实验的方法得到数据后，统计学就用于分析数据。典型的做法包括：制作统计表、制作数据图，或者计算数据的汇总统计量。通常，会用其中的两种或所有的方法。有的方法既适用于定量研究，也适用于定性研究。在定量研究中，为了弥补数据收集方式导致的误差，通常还需用各种方法对数据加权。

定量研究和定性研究的最大差别即在于此。因为数据是随机收集的，所以只有定量研究可以使用统计推断的方法。我们尚不明确，如何在定性研究的基础上进行总体推论。定性研究的结果可以获得事例性的证据，但无论是点估计还是置信区间【Confidence Interval】和 p 值【p Values】都无法使用。然而，通过定性研究累积的洞见常常比定量研究所能得到的结果更深刻。此外，定性研究还可作为后续定量研究的基础。

在计算 p 值和置信区间时，重要的是使用那些可以反映数据收集方法的公式。在实验研究中，通常要计算数据的各种平方和，并用这些平方和计算 F 值，进而得到 p 值。类似的公式并不总是可以方便地应用于更复杂的抽样调查。

如其名称所暗示的那样，定量研究的结果体现为各种量的计算，大多数具有数值特征。对于恰当的随机数据，可以计算相应的汇总统计量，诸如均值和中位数。之后，这些统计量可以帮助我们理解数据中相应的变量。p 值和统计推断的推论的功能也是类似的。

有批评指出，此类汇总的量的信息可能太局限，并不能如实地反映真实的情况。许多统计谬误和统计笑料反映了这点。为了得到对数据的更好理解，即便在定量研究中，也有必要考虑特殊个案的影响，如游离值【Outliers】。当从统计汇总转向个别观测值时，我们就从定量研究转向了定性研究的丰富性。这种对个别数据点的研究，模糊了两种研究方法的界限。一个可能暴露定量研究缺陷的特定领域是对变量间关系的回归【Regression】研究。一个远离观测值集中区域的数据点，可能对回归线和相关系数都有很大的影响。另一方面，诸如中位数这类统计量对游离值并不敏感。

——Gudmund R. Iversen

（马妍译　赵锋校）

参考文献

Bickman, L., & Rog, D. J. (1998). *Handbook of applied social research methods.* Thousand Oaks, CA: Sage.

Brown, S. R., & Melamed, L. E. (1998). *Experimental design and analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-074). Thousand Oaks, CA: Sage.

Kalton, G. (1983). *Introduction to survey sampling* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-035). Beverly Hills, CA: Sage.

定量变量（Quantitative Variable）

统计分析中的变量可以依照渐增的复杂度排列。定量变量最复杂，因为它们使用一定的计量单位来测量。就某一定量变量的任意两个观测而言，它们有相等或不等的值，如用定类变量【Nominal Variable】来测量。同样，它们的取值也可能从低到高排

序，如用定序变量来测量。然而，因为定量变量具有计量单位，两个观测值不仅可能不等，还可以对它们的差异程度予以衡量。定量变量可以是定距变量【Interval Variable】或定比变量。与定类和定序变量相比，更多和更有力的统计方法可用于分析定量变量，回归【Regression】分析即是其中最重要的一种。

定量变量的例子包括年龄、收入和家庭子女数。这类变量的值都有相应的计量单位。年龄的单位通常为岁。以一年为单位，若某人是32岁，另一人是25岁，那么，后者比前者年轻7岁。收入的单位通常是元，也可以使用千元作单位。家庭子女数是一个计数变量，常用单位为个。

然而，任何一个如年龄似的变量，并非必然是定量变量。这一变量的复杂性取决于它的测量方式。年龄可用青年、中年和老年来测量。依此方式，年龄就不是定量变量，而是定序变量。同样，也可用中年和非中年测量它。依此，年龄就是一个定类变量。因此，在使用统计方法时，不仅要知道变量的名称，还必须同时了解变量的测量方式。

定量变量的一个次级特征是它可以是离散的【Discrete】或是连续的【Continuous】。在一个离散变量上，可以找到中间没有值存在的两个取值。计数变量就是离散的；没有任何一个家庭的子女数会在2个孩子和3个孩子之间。类似地，在一个连续变量上，任意两个取值之间总存在着其他的取值。例如，年龄就是一个随时间移动的连续变量，即使这个变量通常以此前一个生日达到的岁数来测量。有时，定量变量也被称为连续变量。

定量变量优于定类变量和定序变量，因为它们可以累加或倍乘，进而为进行多种统计计算和分析提供了可能。定量变量既可以计算均值，也可以进行双变量分析和多变量分析。

——Gudmund R. Iversen

（马妍译　赵锋校）

参考文献

JMP Introductory Guide. (2000). Cary, NC: SAS Institute.

Moore, D. S., & McCabe, G. P. (1998). *Introduction to the practice of statistics*. New York: W. H. Freeman.

SPSS base 10.0 user's guide. (1999). Chicago: SPSS.

四分位数(Quartile)

观测值【Oberservations】根据其排序【Rank Order】可被分为四组规模相等的集。这四组集中的每一个集形成一个四分位数，即一个分位数【Quantile】的特列。

——Tim Futing Liao

（马妍译　赵锋校）

*也可参见分位数【Quantile】。

准实验(Quasi-Experiment)

准实验通过操控假定的原因来探明它们的效应，与此同时，研究者并不以随机的方式将受试单元分入各个实验条件中。由于随机分配并不总是可能的，准实验就是必

需的。伦理的制约不允许在没有恰当知情同意的前提下，根据机会分配处理，不给需要治疗的人以有效的治疗；那些执行实验的人可能拒绝给予随机化分配以应有的重视；或者某一处理已被实施后，人们开始质疑项目的效应；等等。以上诸类问题都会使得随机化变得不可能。因此，准实验将设计特征、实践逻辑和统计分析结合在一起，去证明处理可能是效应的一个可信原因。据此得出的因果推论通常比随机实验条件下的推论更不确定。非随机实验等同于准实验加观测研究（observational study），再加非实验设计（nonexperimental design），通常准实验仅是其子集之一。

准实验设计的种类

准实验设计包括（但不限于）以下种类：(a) 非等价控制组设计。在这种设计中，研究者研究那些暴露于两种或更多种条件的试验结果，但是实验者并不根据实验条件进行控制分配。(b) 中断的时间序列设计【Interrupted Time-Series Designs】。在这种设计中，对某一结果变量予以历时的多次连续观测（以 100 次为基型），并在观测的中段引入处理，从而通过处理后时间序列过程的一次中断来表明处理对结果变量的影响。(c) 回归中断设计（regression discontinuity designs）。在这种设计中，实验者使用某个被测变量上的一个划界分（a cutoff scor）来决定接受处理的资格，同时，如果处理组上分配变量对结果变量的回归线与比较组上的回归线存在断点，就说明存在着处理效应。(d) 单案例设计（single-case designs）。在这种设计中，对一位参与者进行历时的重复观测（通常仅涉及有限的场合，而不构成时间序列），与此同时，操控处理的周期和处理的剂量，由此来证明处理对结果变量有支配作用。

在这类设计中，处理是被操控的，其后结果被观测。还有另外两种设计有时也被包括在准实验中，但是原因项并没有被操控。(e) 在个案控制设计（case control designs）中，将出现了有意思的结果的一组同没有出现有意思的结果的一组相比较，并看一看两组之所以不同，是否是由于一组受到了可能原因的影响，而另一组没有。(f) 在相关设计（correlational designs）中，常通过一次抽样调查，同时对可能的处理和结果进行观测，进而看一看两者是否相关。因为这些设计并不保证原因先于效应，即如因果逻辑所要求的那样，所以通常它们会得出更加不可靠的因果推论。

历史发展

1920 年代前做的大部分实验都是准实验。例如，林德（Lind，1753）描述了针对坏血症的六种医疗处理的一次准实验比较；1850 年前后，流行病学家使用个案控制法发现伦敦霍乱的原因是供应水的污染；而在 1898 年，特里普利特使用了非等价控制组设计表明观众和竞争对手的在场提升了自行车骑手的表现。

1963 年，坎贝尔和斯坦利创造了“准实验”这个词来描述以上类型的设计。坎贝尔和他的同事（Cook & Campbell，1979；Shadish，Cook，& Campbell，2002）从三个方面扩展了这些设计的理论和实践。

第一，他们给出了数量更多的此类设计。例如，有些准实验设计本身即具有纵贯性（例如，时间序列设计和单案例设计），即历时地观测参与者；而有些设计则是通过在处理前或处理后加入更多观测，从而将设计变成纵贯的。类似地，可以使用多个处理组或控制组，也可以将不同的设计结合在一起，比如将一个非等价控制组加入到一个时间序列中。

第二,坎贝尔发展了一种逻辑,而此种逻辑能够评估由准实验设计得出的因果推论的质量——一种在库克和坎贝尔(Cook & Campbell, 1979)与沙迪什等人(Shadish et al., 2002)的著作中得到详细论述的效度【Validity】分类体系。这一分类体系包括四种效度类型,以及对每种效度类型的威胁因素。这些威胁因素即研究者所得因果推论可能是错的常见原因。统计结论的效度涉及推论,在多大程度上假定的原因和效应存在共变关系;而对此构成威胁的一个因素是低的统计检力【Statistical Power】。内在效度【Internal Validity】涉及以下推断,即观测到的共变关系是否可归之于处理对结果的影响;威胁因素之一是历史(即那些同样能引起效应的外来事件)。构念效度【Construct Validity】涉及对研究操作表现出来的更高阶序上的概念的推断;威胁因素之一是实验者期望效应【Experimenter Expectancy Effects】,即参与者以他们认为实验者想要观测到的方式作出反应,而不是对预期的处理作反应。外在效度【External Validity】涉及对如下事项的推断,即随着人群、环境、处理变量和测量变量的变化,因果关系是否仍然存在;威胁因素包括处理同设计的其他特征间的交互作用,因为正是由于交互作用的存在才产生了独特的效应,否则效应就不会被观测到。

第三,坎贝尔使用设计特征来说明效度的威胁因素——即研究者可以操控的诸事项,而通过对这些事项的操控,研究者可以阻止某个威胁因素的出现,或诊断它是否存在,以及它对研究结果的影响(见表1)。例如,人们可以假定,成熟(随着时间的正常发展)即是一个可预料到的效度的威胁因素,因为它可以像处理一样,导致一个前测—后测的变化。在处理前进行几次连续的前测可以表明,处理前的成熟率是否相似于处理期间和处理后的变化率。如果是相似的,那么成熟就是一个威胁因素。所有准实验都是这些设计特征的组合,而研究者需要在特定的背景下,对设计特征作出选择,并以此来诊断或剔除效度的威胁因素。对于用统计调整的方法处理已经发生的威胁因素,坎贝尔持有怀疑的态度,因为统计调整需要作出假定,而这些假定通常是可疑的,或者根本不能检验。

表1 建构准实验的设计要素

分配(控制分配策略以增加组间可比性)

- 划界为基础的分配(Cutoff-Based Assignment)。以一个或多个充分测量的协变量为基础,有控制地将受试者分配入各种实验条件。这种策略会产生一个无偏的效应估计值。
- 其他非随机分配(Other Nonrandom Assignment)。有时近似于随机化分配的各种形式的"偶然"分配(例如,交替将受试者分入一个具有两种条件的准实验中,即将受试者每隔一个单元分入同一条件中,等等。)
- 匹配和分层(Matching and Stratifying)。努力根据观测的协变量,并以稳健且不会导致回归伪造(regression artifacts),又同结果变量有相关关系的方式,创建出等价的群组。

测量(运用测量来了解因果推论的威胁因素是否发生实际作用)

后测观测(Posttest Observations)

- 非等价的因变量(Nonequivalent Dependent Variables)。测量对处理的因力(cause forces)不灵敏,但是对所有或大部分混杂的因力灵敏,而这会导致关于处理效应的虚假结论(如果这类测量显示没有效应,但是结果变量的测量显示出一个效应,因果推论就得到了支持,因为显示出的效应不可能归因于混杂)。

续表

- 多次实质后测(Multiple Substantive Posttests)。用于估计处理是否对那些在理论上可以被预测到,并具有复杂模式的结果存在影响。

前测观测(Pretest Observations)

- 单一前测(Single Pretest)。对结果变量的一个处理前测量,有助于对选择偏倚的诊断。
- 回溯性前测(Retrospective Pretest)。如果实际的前测不具有可行性,就以统计的方式重新构建前测——就其自身而言,这是一种非常不牢靠的设计特征,但有时好过没有前测。
- 代理前测(Proxy Pretest)。如果一个真正的前测不可行,就代之以对某个同结果变量相关的变量实施前测——本身常不可靠。
- 对结果变量的多次时间点前测(Multiple Pretest Time Points on the Outcome)。帮助揭示那些可能把因果推论变得复杂的处理前趋势(pretreatment trends)或回归伪造。
- 对独立样本的前测(Pretests on Independent Samples)。如果不可能对被处理的样本实施前测,那么可代之以对某个可行的等价样本的前测。
- 如预测到的交换作用一类的复合预测项(Complex Predictions Such as Predicted Interaction)。被成功预测到的交互作用能够支持因果推论,因为备择解释变得更不可信。
- 对内在效度的威胁因素的测量(Measurement of Threats to Internal Validity)。帮助诊断对 A 导致 B 这样的推论的特定威胁因素是否存在,如某些受试单元是否积极地寻求实验外的额外处理。

比较组(选择那些“较少非不等价”控制组,或那些可将处理组包括在前测中的控制组)

- 单一非等价组(Single Nonequivalent Groups)。相比于没有控制组的研究,使用一个非等价的控制组则有助于辨识多种效度的潜在威胁因素。
- 多个非等价组(Multiple Nonequivalent Groups)。具有几项功能。如尽可能选出几个同处理组相似的组,但是至少有一组在开始时表现得优于处理组,同时至少有一组的初始表现不如处理组,由此将处理组限定在这些组中。
- 同期群(Cohorts)。比较来自同一机构的不同周期上的群体(例如,同一家庭中的兄弟姐妹或学校里最后一年的学生)。
- 内部(对外部)控制组[Internal (vs. External) Controls]。似乎可以从同一人群中选取控制组(例如,在同一学校内选择,而非选择不同学校的)。

处理(操控处理,以证明处理的变异度影响结果的变异度)

- 移除处理(Removed Treatments)。如果将处理移除,将显示出效应消减。
- 重复处理(Repeated Treatments)。当处理被从某些群组中移除后,重新引入处理——常见于实验室科学或具有短期效应的处理实验中。
- 反复转换(Switching Replications)。颠倒处理组和控制组的角色,即当一组接受处理时,另一组即作为控制组,但是仅当原处理组不再接受进一步的处理或已经被移除处理之后,控制组才接受处理。
- 反向处理(Reversed Treatments)。给予一个概念上相似,但具有相反效应的处理(例如,减少某些学生对计算机接触的同时,增加其他学生对计算机的接触)。
- 剂量变化(Dosage Variation)。用以证明结果变量有规律地对不同水平的处理作出反应。

这个时期也有一些其他学者同样对准实验的因果推论感兴趣,特别是统计学界的威廉·科克伦、经济学界的詹姆斯·赫克曼、流行病学界的奥斯汀·布拉德福德·希尔爵士。当然,坎贝尔的工作是独一无二的,这既由于他更多地强调设计而非统计分析,也缘于他如何评估因果推断的理论,以及过去 40 年间对准实验理论和方法的持久推动。

例子

1966 年,加拿大安大略省发起了一个正式的项目,其目的在于为阻止新生儿发育迟缓,而筛查和治疗患有苯丙酮尿症(PKU)的新生儿。项目开始之后,44 个患有 PKU 的新生儿没有出现发育迟缓,而 3 个新生儿出现发育迟缓。这 3 个新生儿中的两个在筛查项目中没有被检查出来(Webb et al., 1973)。以前的统计数据表明很大比率的发育迟缓是 PKU 造成的。尽管这个研究缺少控制组,作者仍得出结论认为,项目成功治疗了患有 PKU 疾病的婴儿。基于这类实验结果,该项目在加拿大和美国被广泛开展。这个研究就是一个没有控制组的前测—后测准实验。

1982 年 7 月,亚利桑那州开始执行对酒驾严厉惩罚的法令;通过对 1976 年 1 月到 1982 年 7 月每月交通事故致死人数(控制条件)与 1982 年 7 月到 1984 年 5 月每月交通事故致死人数(处理条件)的比较,发现新法生效后交通致死率有所下降。圣地亚哥市于 1982 年 1 月开始执行加利福尼亚州的法律,也开始惩罚酒驾,而该市的月度数据也呈现出类似趋势。以得克萨斯州的埃尔巴索市为一个控制的时间序列——该市在同一时期没有进行驾驶法律上的相关改革,该市的月度致死趋势表明在 1982 年 1 月和 1982 年 7 月两个月没有出现明显的变化。相较于埃尔巴索市的无明显变化,在圣地亚哥市和亚利桑那州发生的历时的趋势变化说明,新的法令降低了交通致死率(West, Hepworth, McCall, & Reich, 1989)。这个研究是一个中断的时间序列准实验。

统计学和准实验设计

统计学家如保罗·霍兰、保罗·罗森鲍姆和唐纳德·鲁宾,既强调研究者要去测量接受处理的参与者若没接受处理会出现何种状况(即反事实),还致力于研究相关的统计方法,即这种统计方法可以改善缺少随机化处理情况下对反事实的估计。一个主要方法是使用倾向值【Propensity Scores】,即通过做实际的群组成员资格对结果的预测变量或影响参与者接受处理的预测变量的 Logistic 回归【Logistic Regression】,获得的一个群组成员资格的预测概率。匹配【Matching】、分层或倾向分上的共变,可以在那些预测项上平衡非等价群组,但是这些方法不可能就未观测到的变量平衡群组,因此,隐藏的偏倚依旧存在。从此以后,这些统计学家已经发展出一些灵敏的分析技术,用以测量隐藏的偏倚会在多大程度上必然严重地改变一种效应。

经济学家詹姆斯·赫克曼和他的同事努力于另一项发展,即选择性偏倚【Selection Bias】模型,其目的在于通过对选择过程建模,从准实验中去除隐藏的偏倚。遗憾的是,在同随机化实验的结果相对照时,这些模型并不十分成功。最近,经济学家通过结合选择偏差模型和倾向分已得出了更好的结果。这一主题还在持续发展。

第三项发展是使用结构方程模型【Structural Equation Modeling, SEM】研究准实验的因果关系;而这一努力只取得了部分成功(Bollen, 1989)。SEM 模型潜变量的能力有时可减少不可信的测量带来的偏倚,但是由于同样缺乏关于选择过程的知识——这一知识上的匮乏妨碍了选择偏倚模型的应用,SEM 还不具备生成无偏的效应估计的能力。一份更新的相关文献提出,可以使用因果方向图(directed graphs)来帮助理解这个问题(Pearl, 2000)。

结论

准实验很难像随机化的实验一样可以

给予因果推论以坚定的信心，因此，在一些领域中过度信赖准实验是一个严重的问题。尽管如此，下列三项重要因素已经为改善准实验创造了条件。第一，基于准实验设计进行的大量实践经验已为研究者提供了一个数据库，从中可以开展对理论的经验研究（Shadish, 2000）。第二，经过几十年对随机设计的倾力研究，统计学家和经济学家已将注意力投向准实验设计的改进。第三，计算机革命同时给予理论家和实践者以更强的能力去发明和应用更加复杂、计算要求更高的方法，以改进准实验设计。

——William R. Shadish

M. H. Clark

（王玥译　赵锋校）

参考文献

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand-McNally.

Cochran, W. G. (1965). The planning of observational studies in human populations. *Journal of the Royal Statistical Society, Series A*, *128*, 134-155.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand-McNally.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*, 153-161.

Hill, A. B. (1953). Observation and experiment. *New England Journal of Medicine*, *248*, 995-1001.

Lind, J. (1753). *A treatise of the scurvy: Of three parts containing an inquiry into the nature, causes and cure of that disease*. Edinburgh, UK: Sands, Murray and Cochran.

Pearl, J. (2000). *Causality, models, reasoning and inference*. New York: Cambridge University Press.

Rosenbaum, P. R. (1995). *Observational studies*. New York: Springer-Verlag.

Schlesselman, J. J. (1982). *Case-control studies: Design, conduct, analysis*. New York: Oxford University Press.

Shadish, W. R. (2000). The empirical program of quasi-experimentation. In L. Bickman (Ed.), *Validity and social experimentation: Donald Campbell's legacy* (pp. 13-35). Thousand Oaks, CA: Sage.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin.

Triplett, N. (1898). The dynamogenic factors in pacemaking and competition. *American Journal of Psychology*, *9*, 507-533.

Webb, J. F., Khazen, R. S., Hanley, W. B., Partington, M. S., Percy, W. J. L., & Rathborn, J. C. (1973). PKU screening—is it worth it? *Canadian Medical Association Journal*, *108*, 328-329.

West, S. G., Hepworth, J. T., McCall, M. A., & Reich, J. W. (1989). An evaluation of Arizona's July 1982 drunk driving law: Effects on the City of Phoenix. *Journal of Applied Social Psychology*, *19*, 1212-1237.

Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, *25*, 659-707.

问卷（Questionnaire）

所有科学家都使用“工具”来测量对他们的学科有意思的现象，社会科学家也不例外。因此，就像望远镜之于天文学家，显微镜之于微生物学家，问卷对多数社会科学家而言是他们的基本测量工具。这些研究者面对的挑战是如何校准问卷，以使其尽可能

精确且高效率地收集数据。

从“总误差”的视角出发，制作不良的问卷可能让收集到的数据出现偏倚【Bias】和较大的方差【Variance】(Groves, 1989)。特别是，糟糕的措辞、糟糕的题序，和/或成问题的问题格式都可能导致明显的测量误差，并表现在偏倚和方差上；同时/或者，就某个题项，产生以偏倚形式表现出来的无回答【Nonresponse】误差。

一份问卷的确定特征之一在于，它是自填式【Self-Administered】调查，如其在邮寄调查或网络调查【Internet Survey】中那样，还是访谈员实施型问卷(interviewer-administered)调查，如其在电话调查或在场调查中那样。如果问卷是自填式的，那么“最终使用者”是回答人。在这种情况下，准确地完成问卷所需的所有填答说明都必须清楚且明确地陈述于问卷之中。或者，在计算机辅助的自填式问卷中，可以通过录音念出填答说明。当然，问卷是自填式的，还是网络调查，还会影响许多其他的设计特征；具体的讨论可以参见迪尔曼(Dillman,2000)的论述。

在创制一条具体的问卷题项(item)时，需考虑四个结构要素：(a)题干(question stem)的措辞；(b)回答选项(response option[s])是开放式的，还是封闭式的；以及如果是封闭式的，那么它们是(c)迫选的(forced-choice)，且/或(d)平衡的。在为题干选择适当的措辞时(即提出的疑问)，最好用尽可能少，且尽可能简单的字词来表达所测概念的意义。同时，在以措辞的方式操作化概念时，研究者也必须注意回答人的教育水平。如果问卷是自填式的，还须考虑回答人的读写能力。

开放式问题容许回答人用他或她自己的语汇来回答问题，从而能就所要测量的内容，给出更加丰富、更加具体的答案，特别是当调查由访谈员来实施时，他可以以无偏见和无引导的方式，探测出更完整的答案(参见 Fowler & Mangione, 1990)。然而，在能系统地分析这些答案前，还需要给开放式回答编码，而要以可靠的方式完成这项工作，这将会是一项既费力又费钱的事。

封闭式回答常表现为有多个选项的样式，同时，这些回答选项给予回答人一系列穷尽(exhaustive)且互斥(mutually exclusive)的选择，并请他们从中选出一个。如果调查由访谈员实施，有时封闭式问题本身就隐含着回答选项，如那些“是或否”之类的问题。有些问题类型还可以有多个答案(例如，问人们喜欢的三种休闲活动)。那些对回答人的认知过程——一个回答人用以感知一个调查问题，然后决定如何作答的心理过程——感兴趣的读者，可以参见图朗若、里普斯和拉辛斯基(Tourangeau, Rips, & Rasinski, 2000)的著作。

在为多数封闭式题项设计回答选项时，研究者必须考虑，这一选项集应否包含一个“中性点”和/或“不确定”(如不知道)的回答项。如果回答选项集不容许回答人使用一个中性点(例如，下列回答选项集中的中间项：同意、既不同意也不反对、不同意)，或者不让回答人表明他或她是“不确定”的，这样的题项就被称为迫选题。非迫选题给予回答人一个清晰的“在……之外”的选择，如回答人可以简单地表明他或她是“不确定”的。选择何种方式去措辞一个调查题项应考虑：(a)要测量的现象，以及(b)回答人给出包含细微差别的答案的能力和意愿(关于如何“满足”回答人的讨论，参见 Krosnick, 1999)。

问题设计中的另一个结构要素的考虑是，回答的量尺是平衡的还是不平衡的。一个平衡的回答选项集有一个真正的中点，不管这个中点本身是否作为一个选项被给出。以此，强烈同意、同意、不同意和强烈不同意这一组回答选项就是平衡的，即便其中没有明白地给出一个真正的中点。相反，如极

好、较好、一般、较差，这样的回答选项集就是不平衡的，因为其中有两个选项是正向的，有一个是中立的，还有一个是负向的。

问卷中问题的题序或问题的题区可以影响所得数据（Schuman & Presser, 1981）。一般而言，问卷的组织最好是从容易回答、可以建立良好关系的题项开始，即使这些题项可能不会用于分析。这样做会降低回答人在开始填答后，又马上拒绝完成问卷的概率。后面可以摆放测量观点和态度的题项，再随之以测量行为和知识的问题。这样做的原因在于，相对于观点和态度的问题，对行为和知识问题的回答据认为较少受题序的影响。问卷应以个人背景信息和人口学信息做结尾。一份问卷的内容如此安排的理由在于，就极小化无回答可能性而言，这样做最恰当，同时，也使得所有类型的题项都不太可能影响后续问题的作答。

强烈建议，在收集实际的研究数据之前，对所有的问卷实施试调查。试调查能给出一些很重要的反馈信息，如回答人对问题措辞理解的程度、问卷的长度，以及回答人的负担。在许多时候，聪明的做法是进行一组认知访谈，即同试调查回答人进行一对一的“出声思考（think aloud）”会谈，其中回答人要清楚地讲出他们在努力理解和回答每一个问题时进行的思考过程。

最后，在设计问卷时，研究者还应当有一个针对数据减损和如何处理缺失数据【Missing Data】的事前计划。数据减损包括用几个不同题项的数据生成新变量的过程中产生的数据损耗，如将几个量表【Scale】合并在一起时出现的数据损耗。而缺失数据可以用几种方式来处理，包括用统计的办法给每一个缺失数据赋值。

——Paul J. Lavrakas

（王玥译　赵锋校）

参考文献

Dillman, D. A. (2000). *Mail and Internet surveys: The tailored design method*. New York: Wiley.

Fowler, F. J., Jr., & Mangione, T. W. (1990). *Standardized survey interviewing*. Newbury Park, CA: Sage.

Groves, R. M. (1989). *Survey errors and survey costs*. New York: Wiley.

Krosnick, J. A. (1999). Survey methodology. *Annual Review of Psychology*, *50*, 537-567.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys*. New York: Academic Press.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, UK: Cambridge University Press.

排队论（Queueing Theory）

社会科学家常研究这样一种情境：存在着对有限资源的过度需求。排队论是随机的【Stochastic】建模的一个分支，也是一个假定对特定服务或资源有需求的人等待协调的过程。许多文献都广泛使用这种方法来描述完成时间不确定的流水线式的过程和服务。该理论最初用来分析电话占线和工业吞吐量（Brockmeyer, Halstrom, & Jensen, 1960; Erlang, 1935），它利用一组描述供给的三个方面的规则来解释结果：需求者进入服务系统的方式；供给者的供给能力和资源量；个体相对于其他个体被对待的方式。

模型的第一部分规定：业务根据假定的

参数形式随机到达。如果我们可以声称业务的到达是独立的和顺序的（不同时到达），那么排队论模型是非常容易描述的。它并不要求这些业务在接受服务以前，都驻留在分配的或选定的队列中；它们可能因为队列长度而立即离开（终止）；或者在等待了一段时间仍未被服务后离开（放弃）；或者可能换到系统里另一个有空闲的队列（移动）；参数化的到达率可能随时间指标变化（不平稳），或者随时间固定不变（平稳）。此外，业务可以被赋予不同的特性。这些特性可以用来表示优先级、复杂性或标志某行为曾在系统中出现过。

模型的第二个主要部分是对系统服务资源的描述。其属性包括服务队列（或渠道）的数量、优先队列存在与否、服务要求的层级，以及缓冲或存储的容量。这些特征及其安排方式决定了系统处理待服务业务的效能。

模型的第三个也是最后一个部分是一组管理队列纪律的规则。它们包括排队业务的服务顺序、优先级的安排，以及对潜在"插队"的容许度。两种最常见的业务服务安排是"先到先出"和"先到后出"。此外，优先级的安排可以界定为当重要业务进入系统后马上对其进行服务，即通过取代正在接受服务的业务，同时将非特权优先的个案置于队列的前端，使其成为下一个被服务的对象。

Kendall 记号法（以大卫·肯德尔爵士命名的）描述了任何给定的排队论模型的假设。其形式为：$A/B/\sigma/k/M/Z$。其中，A 表示到达时间的分布；B 表示服务时间的分布；σ 表示服务者的数量；k 表示队列中待服务或可容纳的业务总量；M 表示业务来源的总体规模；Z 表示对描述排队纪律的描述。

例如，$M/D/3/8/40\ 000/FIFO$ 描述了一家理发店的服务：M = 顾客到达时间呈指数分布，D = 确定型的服务（同时减少），3 = 3 个理发师，8 = 有 5 个位置给等候的顾客，40 000 = 在一个中等规模的小镇，FIFO = 按照到达顺序进行服务。然而，常规的做法是：删去最后 3 项假设，若根据问题的背景，它们很明显，特别是总体规模，因为通常它既无法控制，也不影响模型设定。

最简单和通用的模型设定是 $M/M/\sigma$，指数分布的到达时间间隔（因此，泊松分布的达到时间间隔是 t），指数分布的服务时间，并且有 σ 个服务者。大多数教材（Gross & Harris，1998）都始于该设定，因为，使用宽泛的假设，许多共同的经验情境才能被有效地描述。

——Jeff Gill

（马妍译　赵锋校）

参考文献

Brockmeyer, E., Halstrom, H. L., & Jensen, A. (1960). *The life and works of A. K. Erlang.* Kobenhavn: Akademiet for de Tekniske Videnskaber.

Erlang, A. K. (1935). *Fircifrede logaritmetavler og andre regnetavler til brug ved undervisning og i praksis.* Kobenhavn: G. E. C. Gads.

Gross, D., & Harris, C. M. (1998). *Fundamentals of queueing theory* (3rd ed.). New York: Wiley.

配额样本（Quota Sample）

配额样本是一种为调查选取回答人的方法。调查实施者分配给访谈员一定的配额，让他们根据一定量关键的人口学特征来确定回答人的组别。访谈员则选取那些同所给特

征相匹配的个体作为回答人,进而完成配额。当所有回答人的回答被结合在一起时,样本的特征精确地同具体调查总体的人口学特征相匹配。例如,一名访谈员可能被分给如下配额,即样本中包含四名非裔美国人、八名白人、两名西班牙裔美国人,还有两名亚裔美国人;同时,他们中的一半生活在中心城区,另一半生活在郊区。当所有访谈完成后,就作为访谈标准的人口学特征而言,样本【Sample】的总和统计数和研究总体【Population】的总和统计数将完全一致。然而,总和统计数上的一致并不保证样本会在研究感兴趣的变量上准确地反映出总体特征。

出现在配额抽样中的最令人诟病的偏差【Bias】是1948年为总统选举进行的民意调查。这次选举不利于杜鲁门,他是民主党的候选人,且在罗斯福总统去世时被指定为总统;他的对手是共和党候选人托马斯·杜威。三个民意调查公司(Roper, Gallup 和 Crossley)根据用配额抽样方法开展的访谈调查,都宣布杜威是赢家。当然,杜鲁门最终以高于杜威近5个百分点的选票击败了杜威,并以略低于50%的选票赢得大选。虽然配额样本已经很仔细地在年龄、种族、性别、租金支出、居住条件诸方面同普查数据相匹配,访谈员却选取了过多的共和党选民。共和党选民更容易接受访谈,这或者是因为他们更愿意参加访谈,或者因为他们更容易接触,于是,在访谈中,民主党选民的代表人数则不够充分。选举后,报纸中充斥了关于民意公司和许多广播公司失败的故事。随后,三家民意调查公司都放弃了配额抽样技术。

这一事件(没能准确预测1948年大选结果)推动了概率样本【Probability Sample】的使用,还极大地限制了配额样本在社会科学研究和民意调查中的应用。目前,大多数选前民意调查依赖于随机数字拨号【Random Digit Dialing, RDD】来选择家庭,再从中随机选取成年回答人,或者依赖于从一份自动生成的资格选民的清单中,由计算机随机生成一份选中的样本。具体的个体由这些调查程序予以辨识,并且不容许访谈员替换回答人,进而杜绝了访谈员的自行决断和选择偏差。

偏差和配额抽样

配额样本之所以不能提供关于总体的准确信息,其主要原因在于,访谈员拥有访谈谁的自行决断权。无论何时只要容许人为判断能决定谁被选中,即便这时访谈员必须满足人口学配额,无意识的偏差总会溜进选择过程。虽然偏差是无意识的,它仍能在研究数据和数据生成的估计值(例如,均值、比例或回归系数【Regression Coefficient】)上引起显著的误差。

配额样本是非概率样本(nonprobability samples),有时也被称作目的【Purposive】或方便抽样【Convenience Sampling】。非概率样本的共同特征在于,人为判断在决定选择哪一个人来回答访谈员的问题时发挥作用。非概率样本常被当作概率样本(probability samples)的反面,而后者则要求,研究总体中的每个个体都有一个(已知的)非零的概率被选作回答人【Respondent】。概率样本依赖某种随机机制从研究总体中找出向访谈员作答的个体成员。回答人一旦被找出,就不容许访谈员替换回答人(Hess, 1985)。如果找到的个体不能接触到或拒绝回答,这个个案就被视作缺失数据,同时将其计入研究的无回答偏倚【Nonresponse Bias】。若使用配额抽样,一次拒绝就会导致访谈员去寻找另一名具有一定特征的个体,再访谈新找到的人,以填补他或她的配额。由于配额样本既不始于研究总体的清单,也不保留拒绝调查的记录,这就不可能去分析其最终样本中的潜在偏差。不过,人们已经进行了许多

研究,并发展了一些分析程序来估计由概率样本调查中的无回答产生的偏差(Keeter, Miller, Kohut, Groves, & Presser, 2000)。

与配额抽样相关的抽样方法

尽管配额样本在当代社会科学中使用得相当少,有几种相似的抽样方法却常被使用。有一种概率抽样方法,即比例分层抽样【Stratified Sampling】(Henry, 1990; Kish, 1965),有时会被混同于配额抽样,因为两者都可以精确地复制研究总体的一些人口学特征。不过,比例分层抽样使用一种随机选择程序,而不是让访谈员去选择回答人。要使用比例分层抽样,就要得到一份研究总体的清单,其中还要包含有用于进行匹配【Matching】的变量的信息。而后,研究者根据关键变量,将所有个体分进互斥的组或层。研究的抽样分数即被抽作样本的人数占总体的百分比($f = n/N$,其中,f 是抽样分数【Sampling Fraction】;n 是样本量;N 是总体规模)——被施用于每一个层,并由此确定每一个层的样本量($n_s = fN_s$)。最后,通过一种随机过程从每一个层中选取所需的回答人(n_s)。这一过程可以确保,所选样本和研究总体在关键变量上保持一致,同时摒除了作为一种偏差的访谈员选择。

为了利用配额样本的快速和低成本的优点,有些调查研究者试图将配额抽样和比例分层抽样结合到一种被称作"配额概率抽样"的技术中(Sudman, 1976)。这一技术将访谈员分入特定的邻里社区中去实施访谈,通常是某个特定时段上的某一特定的街区,同时,要求他们根据包含了该特定街区的普查区的若干人口学特征进行样本匹配。通过加强对访谈员的控制和记录保存,就可能估计和减少由拒绝参访带来的偏差。

另一种在抽样中使用配额的方式是,在一个分层的样本中,为来自每一层的访谈数量设定限制。通常,这样做是为了给予那些被过低代表的群体以更多的样本量,如少数族裔。所有回答人都通过概率抽样的方法选出,但是一旦某一层的配额达到限度,在获得了能识别它们是该层成员的信息后,针对该层成员的访谈即终止。

有时,定性研究和其他研究,为了从研究总体中获取一个异质群体的数据,也会使用配额。尽管这些研究会产生有偏的总体估计,它们却能增加将对立案例纳入研究的可能性,而这些案例间的差异可通过研究得到探查。

——Gary T. Henry

(王玥译　赵锋校)

参考文献

Henry, G. T. (1990). *Practical sampling*. Newbury Park, CA: Sage.

Hess, I. (1985). *Sampling for social research surveys 1947-1980*. Ann Arbor, MI: Institute for Social Research.

Keeter, S., Miller, C., Kohut, A., Groves, R. M., & Presser, S. (2000). Consequences of reducing nonresponse in a national telephone survey. *Public Opinion Quarterly*, *64*, 125-148.

Kish, L. (1965). *Survey sampling*. New York: Wiley.

Sudman, S. (1976). *Applied sampling*. New York: Academic Press.

R

R

R 是多元相关系数【Multiple Correlation Coefficient】的符号。双变量相关系数是反映两个变量（如 X 和 Y）之间的相关程度。多元相关系数反映了两个或多个自变量【Independent Variable】（如 X_1，X_2 和 X_3）同因变量【Dependent Variable】Y 之间的相关程度。在这一例子中，该系数将告诉我们三个自变量如何共同同因变量 Y 线性地相关。R 的理论值介于 0～1。大的取值表示高度线性相关，小的取值表示低度线性相关，但这并不排除高度的非线性【Nonlinearity】。需要注意的是，R 也用作一个统计软件的名称，与 S-plus 有许多类似性，且在社会科学领域日益普及。

——Michael S. Lewis-Beck

（马妍译　赵锋校）

r

r 是皮尔逊相关系数【Pearson's Correlation Coefficient】的符号，由样本【Sample】予以估计。重要的是要将它与总体参数 rho（ρ）区别开。它是测量两个定量变量【Quantitative Variable】之间相关度的主要指标，可用来估计线性关系的强度，取值范围为-1～+1。

——Michael S. Lewis-Beck

（马妍译　赵锋校）

随机分配（Random Assignment）

随机分配是指将研究对象分配到某一实验【Experiment】中的各试验组的方式。一个试验组是典型的控制组【Control Group】，还可以有一个或多个实验组。这个方法能够抵消所有其他变量的系统效应，从而更好地估计处理【Treatment】变量的效应。随机分配使得应用统计推论方法获得数据所来自的总体的结论成为可能。

许多关于如何做实验的主要观念源自英国统计学家罗纳德·费希尔爵士。他做了大量的农业实验。在用许多块田块做实验时，问题在于人们如何将不同的田块分入控制组和实验组，同时使得其他非处理的变量的影响最小。他希望分配能够做到，不存

在能够解释实验结果的潜在因素。除了实验变量,所有其他变量的效应应该抵消。虽然这些变量的效应不可能完全消除,但是它们的效应不应系统地存在于数据中。无论所有其他变量的效应是什么,它们都将被归入残余变量中。

费希尔对这个问题的回答是,以随机的方式把不同的要素(田块、人、动物、植物等)分配到不同的处理中。随机分配意味着将要素分入不同的处理组,而这些处理组除了实验变量的影响外,还存在所有其他变量的系统性的影响。当只有一个控制组和一个实验组时,实现随机分配的一个方法是为每种要素投掷硬币。出现"背面"就将某个要素分入控制组,出现"头面"则分入实验组。完成分配后,在每一组中应有数量相同的要素,同时,也不会存在对实验结果的所有其他变量的系统影响。

掷硬币是一种相当原始的将要素分入处理组中的方法。另一种是使用随机数表【Random Numbers Table】。选定一个随机起点后,人们可以使用奇数或偶数。一些计算机软件包生成随机数码,据此,人们可以把要素随机地分入控制组和实验组。

当不是只有一个实验组时,也能进行要素的随机分配。如果费希尔打算比较四种不同类型的谷物,要依类似的随机方式,将田块分入四个实验组和一个控制组中。

应用随机分配的目的与观测研究中的随机抽样【Sampling】方法相同。之所以要随机分配和随机抽样是为了实施统计假设检验【Hypothesis Testing】和进行置信区间【Confidence Interval】的构建。

——Gudmund R. Iversen

(王玥译 赵锋校)

参考文献

Brown, S. R., & Melamed, L. E. (1998). *Experimental design and analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-074). Thousand Oaks, CA: Sage.

Cobb, G. W. (1998). *Introduction to design and analysis of experiments*. New York: Springer-Verlag.

Fisher, R. A. (1935). *The design of experiments* (7th ed.). New York: Hafner.

随机数字拨号(Random Digit Dialing)

随机数字拨号是用于电话调查【Telephone Surveys】的一种抽样技术。代之以使用可能不完整的或过时的电话号码簿,研究者使用随机数字来组成电话号码。目标地区的数字代码不是随机拨通的,而是简单地放在随机生成的数码的起始位置。在美国,这一般意味着将一个随机生成的四位数码加到一个三位的地区交换代码之后。

——Alan Bryman

(王玥译 赵锋校)

随机误差(Random Error)

随机性是统计推论【Statistical Inference】的核心。随机性基于如下观念:如果重复一项观测或测量,那么在绝大多数情况下,将得到一个不同的观测值。

在一次调查【Survey】或实验【Experiment】中,若不可能预测下一次观测的结果,即存在随机性。因此可以认为,如果一个变量具有随机性,那么观测值中会呈现随机变差【Random Variations】,该变量称为随机变量。随机性的大小由随机误差来衡量。

随机变量观测值的变化会引起由数据算得的所有样本【Sample】统计量的变化。此即,一个样本均值【Mean】随样本的变化而变化,如同支持某一候选人的投票者的比例会随投票群体的变化而变化。

若抽取多个样本,从一个样本到另一个样本的样本统计量的变化就会显现。一个专门的民调机构不会重复进行抽样,但在任何一个重大选举之前,几个民调机构会开展自己的研究,他们报告的比例则各不相同。差异的原因之一在于每一抽样过程中固有的随机性。我们可以用适当的统计软件模拟此类变化。例如,模拟抽取样本量 $n = 1\,000$,并以 0.5 的概率从“1”和“0”两个数字中随机取样,那么最初几个样本中抽得“1”的比例可能是:50.4%,50.4%,48.4%和 51.4%。

此类变异即所谓的与比例相关的随机误差。这类变化本身虽无任何“错误”可言,可是不幸的是,“误差【Error】”一词同这类变化联系在一起。不过,就某种程度而言,由于总体中存在一个真实的比例值,那么任何一个样本比例值相对于此一真实总体【Population】参数的任何偏离都可被视为一项误差。

历史说明

物理学家一直都在考虑如何测量物理对象的特征。即使我们清楚地知道一把码尺(给定固定的问题等)只有唯一的、固定的长度,对该码尺长度的重复测量也会得到不同的结果。最终,物理学家接受了这种变异,并用“误差”一词来命名一个观测值与真值(潜在值)之间的差异。当观察到一个统计量由于不同的样本间而有不同的值时,统计学家继承了这一术语,不过,此类差异则源于抽样效应。

应用

随机误差的存在引出关于此类误差的大小问题。若随机误差的大小是已知的,那么人们就可以说某个特定的样本统计量是否超过了随机误差。还可预测一个样本的百分比在真实参数值周围多大范围内变动。统计学家发展了许多计算随机误差大小的公式。如果这些公式不存在,通常可用适当的软件模拟出多次重复的样本,再计算随机误差的大小。

例子

测量随机误差大小的一种方法是获得某个样本统计量在多个不同样本中的标准差【Standard Deviation】。这一标准差通常被称为样本统计量的标准误【Standard Error】,以此区别从一组单个观测值计算得到的标准差。假设从一个 50 对 50 的总体中抽取一个规模为 1 000 的样本,得到样本比例的标准误为 1.6。在大多数情况下,两倍的标准差或者两倍的标准误能包括样本比例值的绝大部分。$2\times1.6=3.2$,因此,大多数从各个样本算得的样本比例值,将介于 46.8($50-3.2=46.8$)和 53.2($50+3.2=53.2$)之间。

然而,最令我们感兴趣的是,某个特定的样本统计量是否位于抽样过程产生的随机误差范围内,或者是否样本统计量远离推定的总体参数值。设想一下,假设一个总体的比值等于 50%。根据一个样本量为1 000的样本观测值,算得的样本比值为56.7%。样本比值 56.7%位于总体参数值 50%的随机误差范围内吗?抑或,由于抽样程序的问题,56.7%超出了随机误差的范围吗?

现在可以得出如下结论,即样本比例值超出了抽样误差的范围。可以将样本比例值换算为一个标准正态变量值。若该值大于1.96,我们则推论,样本统计量同假定的总体参数值的偏差超出了基于随机误差所能期待的差值。这里,$z=(56.7-50.0)/1.58=4.24$。在10 000个不同的样本中,也未必能得到一个样本,其样本统计量的标准正态变量值有如此之大或更大。由此,我们进而推论,总体中有50%的人支持候选人的假定是错误的。因为观测到的样本值应位于总体比例值的随机误差范围内,所以总体比例值必定位于接近样本观测值56.7%附近。

任何样本统计量都有其随机误差,即它的标准误。可通过选定样本收集的方式和特定研究中所需要的变量来控制随机误差的量值。减小随机误差的一个方法是抽取一个更大规模的样本。样本中观测值的数目通常直接进入样本统计量标准误的计算。它的一个缺点是样本量常以平方根的形式出现在计算公式中,因此需要大得多的样本量来获得预期的效应。由于是平方根,所以就需要将样本量扩大四倍,才能得到比原来小一半的标准误。在多元回归【Multiple Regressions】中,共线性【Collinearity】会增大回归系数的标准误。

——Gudmund R. Iversen

(马妍译　赵锋译)

参考文献

Agresti, A., & Finley, B. (1997). *Statistical methods for the social sciences* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Iversen, G., & Gergen, M. (1997). *Statistics: The conceptual approach*. New York: Springer-Verlag.

Moore, D. S., & McCabe, G. P. (1998). *Introduction to the practice of statistics*. New York: W. H. Freeman.

随机因子(Random Factor)

因子是变量的另一种说法。因子这一术语常用于对实验【Experiment】数据的分析中。在实验中,"干预"一词也用来代替因子或变量。对观测数据而言,干预【Treatment】通常不是一个恰当的术语。在宗教归属研究中,对天主教、犹太教、新教、其他宗教这一定类变量而言,干预不是一个好的术语。因为,不能像根据施肥量划分地块一样,人为地将人们划入不同的宗教团体中。

然而,如果一项实验研究,欲考察某个定类变量【Nominal Variable】对一个定量变量【Quantitative Variable】的效应,那么术语"干预"和"因子"是可以互换的,而且这两个术语都很常用。此类研究可称为方差分析【Analysis of Variance】,由于它同实验研究的历史联系,干预或因子两术语常用于代替更中性的术语——变量。在此类研究中,自变量【Independent Variable】(干预、因子)具有不同的水平,而非有不同的值。

方差分析中的因子可以是两类之一:固定的(fixed)或者随机的(random)。我们使用一个因子的所有可能类别,该因子就被视作固定的。如性别,当我们用男和女作类别时,就是一个固定因子。就大多数研究目的而言,此变量再无其他类别。一个因子若我们只有它的几个类别的数据,则是随机因子。在一项关于50个州的差异研究中,首先选出一个随机样本,比如说,10个州,再从每个被选的州中收集因变量【Dependent Variable】的数据。

随机因子是一类典型的变量,带有大

量的类别,以至于对每个类别收集因变量的数据,或者不可能,或者不经济。于是,首先从所有类别中选取一个随机样本,再将研究限制在这些类别中,研究才成为可行的。

以固定因子作为自变量的方差分析可记作方差分析模型I【Model I ANOVA】;类此,使用随机因子的分析记作方差分析模型II【Model II ANOVA】。在仅有一个因子的情形中,两类模型计算出的平方和与 F 值【F Values】均相同。但若含有两个或更多的因子,无论使用模型 I 还是模型 II,平方和与 F 值的计算结果都不同。例如,在一个双向方差分析中,用模型 I 作分析,用行、列和交互项的均平方去除残差的均平方,以获得 p 值【p Values】。但用模型 II 作分析,则用行和列的均平方去除交互项的均平方,来获得 p 值。

分析中也可包含混合因子,一部分是固定的,一部分是随机的。因为有必要区分一个因子是固定的还是随机的,在分析数据时,人们就应当确定相关的统计软件中是否已经考虑到此项区别。

——Gudmund R. Iversen

(马妍译　赵锋校)

参考文献

Agresti, A., & Finley, B. (1997). *Statistical methods for the social sciences* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Box, G. E. P., Hunter, W. G., & Hunter, J. S. (1978). *Statistics for experimenters.* New York: Wiley Interscience.

Jackson, S., & Brashers, D. E. (1994). *Random Factors in ANOVA* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-098). Thousand Oaks, CA: Sage.

随机数生成器(Random Number Generator)

随机数生成器是产生一组无法预测的数值序列的程序。随机数生成器广泛地用于模拟【Simulation】、蒙特卡罗统计技术和密码技术。真实的随机数产生于自然过程,如放射性衰变或天电噪声。然而,大多数应用使用虚拟随机数。它们是由一个决定方程生成的一组具有类似于真实随机序列的统计属性的数值序列组成的。虚拟随机数生成器已被纳入多数计算机语言和数值软件包之中,如电子表格和统计软件包。

在生成虚拟随机数的生成器中,最常用的技术是线性同余发生器:

$$y_{n+1} = (ay_n + b) \bmod m$$

例如,在 ANSI C 计算机语言中,rand( ) 作为随机数函数,有 $m=2^{31}$,$a=1\ 103\ 515\ 245$,$b=12\ 345$。出于计算效率的考虑,m 通常是 2 的幂;a 和 b 基于该方程所生成的序列特征,试验地选取。线性同余函数产生 $[0,m-1]$ 间的整数;若 y_n 被 $m-1$ 除,于是可以得到 $[0,1]$ 区间的均匀分布的随机数列。均匀分布数列进而可用于产生满足其他概率【Probability】分布的随机数列,如使用专门的算法【Algorithms】或倒转合适的累积概率分布函数生成正态分布随机数列。

因为线性同余发生器只用 y_n 决定 y_{n+1},所以一旦某个数值不止一次出现,其后序列中的后序数值即以该数值第一次出现后相同的顺序重现。生成的非重复的数值的数量被称为生成器的“循环”或“周期”。这个数必然小于 m,在造得不好的生成器中,这

个周期可能相当短。

一个备选的方法称为滞后斐波那契生成器(the lagged Fibonacci generator),其形式如下:

$$y_{n+1} = (y_{n-i} + y_{n-j}) \bmod m$$

就某些正整数 i 和 j,斐波那契生成器可以有相当长的周期,一般为 $10^{200} \sim 10^{500}$。

用在某个线性生成器的初始数值被称为"种子",依此数值可以产生不同的序列。在某些应用中,相同的种子被重复使用,以便每次有相同的序列产生;在另一些应用中,常用某些随机变化的物理过程,随机设置种子,如计算机开启后的秒数。

发展高质量的虚拟随机数生成器被证明是计算机科学中相当困难的问题;努斯(Knuth,1997)在此问题上作出了权威性的工作。用一组统计检验可以估计生成器的质量;高质量生成器的关键特征是长周期,y_n 和 y_{n-k} 间缺乏序列相关,且均匀地填充在区间[0,m-1]之间。最后,因为计算应用常常需要有数以万亿的虚拟随机数的产生,所以计算效率很重要。

真实随机数常用于专门的计算应用,如密码技术,其中,序列中的任何模式都可能弱化整个系统。一些网页提供了基于物理过程的真实随机数,如用盖革计数器(Geiger counter)检测放射性衰变,参见相关网站。无理数中小数位位数,如 π 或 2 的平方根也可形成真实随机序列。网站相关内容给出了来自各种无理数的若干百万位的随机数序列。不同于虚拟随机数,真实随机数序列从不循环。

——Philip A. Schrodt

(赵锋译校)

参考文献

Knuth, D. E. (1997). *The art of computer programming, volume 2: Seminumerical algorithms.* Reading, MA: Addison-Wesley.

随机数表(Random Number Table)

我们都见过为了确定谁是奖项的赢家而从帽子中抽取一个或多个号码的情形。例如,先将入场券的票根置入一个碗中摇晃,随机抽出一个票根,读出中奖号码。事实上,这一功能可由随机数表完成。表 1 简要地摘录了某随机数表中一页的两行数字(五位一组的数之间的间隙只是为了方便阅读,因此,在翻页阅读随机数表时,应当忽略)。

表 1 从随机数表中摘录

77 921	06 907	11 008	42 751	27 756	53 498	18 602	70 659	90 655	15 053	21 916	81 825	44 394	42 880	
99 562	72 905	56 420	69 994	98 872	31 016	71 194	18 738	44 013	48 840	63 213	21 069	10 634	12 952	

随机数表中的数字要满足两个基本属性:

1.任一数位上的数字从 0~9 应有相同出现的概率(出现的等可能性)。

2.任一给定数的概率不受其先行和后续数值的影响(独立性)。

抽出一组适于统计分析的随机样本是随机数表的一个主要用途。假设有一个由 N 项构成的总体,且希望选出一个由 n 项组成的随机抽样【Random Sampling】样本。我们首先分配一组数码给 N 项总体中的每一个元素,如整数 0,1,…,N-1。从随机数表读取的数码与这些数码对应的那些总体中的项将构成样本的元素。步骤一,在表中选取一个随机起点。例如,你可以闭上眼,随意翻开书的一页,用手指点中一个起始数。以此数为起点,可以从左至右(或从右至左,或从上至下,或对角,等等)一个接一个地读取。如果 N-1 有 x 位,我们依次相继读取 x 位数码,并忽略任一数字间的间隙。

例如,假设我们有 683 项的总体,以上面表格的摘录作随机起点,从表格的左上角开始。由此,从左至右读数,包含于样本的项目数码依次是 779,210,690,711,008,427,512,775,653,498,以此类推。同时,必须滤去大于 683 的数码。我们持续依次读取三位一组的数码直至获得 n 个选项。样本既可以重复选取(如果总体项目号 328 不止一次出现在随机数码清单上,那么该选项就不止一次入选样本),也可以无重复选取(如果项目号 328 出现超过一次,第一次选取后的重复都被滤除)。

重要的是在每次使用随机数表时,要确保从表中不同的起始点开始。有时,人们希望用随机数表本身选取起始点。例如,我们可以从上面的节录中读取数值 56,表示我们应当从第 5 页的第 6 位开始。

第一张随机数表产生于 1947 年。1995 年,RAND 公司出版了一本题为《百万随机数》(*A Million Random Digits*)的书;该书仍在售,也可以登录相关网址查阅其数字版。2002 年该书还发行了平装本。

——Jean D. Gibbons

(赵锋译校)

随机抽样(Random Sampling)

随机抽样是指能够满足必要标准,从而容许应用以随机化为基础的统计推断【Statistical Inference】的调查抽样设计。术语概率抽样【Probability Sampling】是其同义词。一种随机抽样设计的确切特征如下:

1.对研究总体有全面的界定。

2.设计是详定的。

3.设计是客观的。

4.设计是可重复的。

5.总体中的每一个单元都有一个已知的被选中的概率。所有的选取概率都是非零的。

这些标准还需进一步说明。第一个标准似乎很明白,却常被忽视。必须精确地界定调查的兴趣单元和总体边界。例如,“所有成年人”这种说法是不充分的。定义中应包含地理区域、时间、公民身份、年龄范围及其他相关因素。更好的定义可能是“所有 18 岁及以上,在 2003 年 7 月 10 日,他们唯一或主要的居所是位于美国 48 个大陆州之一的一所私人住宅”。如果调查在一段较长时间内开展,我们还应说明,如何处理常住人口和流动人口间的区别。

详定(标准 2)必须对选取样本的过程有一个清楚且全面的描述。

客观性要求每一次选取——以及选取过程的每一个阶段(由于一些设计是多阶段的)——都由某种随机机制来控制。不允许任何能影响样本单位选取的主观判断。这可能是随机抽样定义的核心。许多研究显示,无论是专家还是非专家,如果请他们有目的地选出一个他们认为有代表性的样本,

他们得到的样本都会在某些重要的特征上存在严重误差。教抽样的老师在课堂练习中也经常观察到这一现象。利用某种机会机制选取样本是唯一可以避免系统误差【Systematic Error】(同时也可以衡量变量误差)的方法。

可重复性【Replicability】要求,在给定同样条件下,另外的研究者能够照样实施同一设计。这并不意味着他或她必须再次得到选中的样本(实际上,这是极不可能发生的),而仅指他或她能重现抽样设计,也就是再现抽样过程的诸阶段,并在恰当的阶段再次应用同样的随机机制。

标准5表明,有必要知道总体中每个单元的选中概率。实际上,从做估计这一目的出发,仅需要知道那些实际上被选中的单元的概率——不必包括未选中的单元。这一标准的关键在于,在原则上,必须能够计算出总体中任一单元的选中概率。要满足这一要求,首先就必须有一个详定且客观的设计,此外,还必须在样本选取的整个过程中收集和记录必要的信息。例如,如果设计涉及从一份清单中选取居住地址,随后从每一个选中的地址中随机选出一个人来访问,就必须记下每一个样本地址内符合条件的人数。

最后的标准是总体的每一个单元必须有一个被选中的机会。不允许将任何一个单元完全排除于选择过程。如果出现了这种情况,同时,被排除的单元以某种系统的方式不同于被包含的单元,就会发生"覆盖偏差"(coverage bias)。

随机抽样被广泛地认为是从中可以产生描述性统计量的公共部门暨学术定量调查唯一可接受的方法。当不存在其他来源的系统误差(如无回答偏差【Nonresponse Bias】和测量偏差)时,它可以为无偏【Unbiased】估计提供基础。然而,基于模型的推论【Inference】方法认为,随机化不必然保证,样本结构中的系统误差源能够在模型中被反映出来(如作为模型的协变量)。由此逐渐增多的做法是,研究者以随机抽样为基础,通过使用模型辅助的推论,将两种方法结合起来(Brewer, 2002)。

许多政府和学术性社会调查都要求随机抽样。在一些协议中,如欧洲劳动力调查(the European Labour Force Surveys)(Eurostat, 1998)协议、欧洲社会调查(European Social Survey)(ESS-CCT, 2001)协议,以及国际社会调查项目(International Social Survey Programme)(Park & Jowell, 1997)协议,对这点有清楚的规定。然而,人们常常发现,有些声称使用随机抽样的调查实际上在一些重要方面背离上述标准——可能允许在第一阶段有目的【Purposive】地选取样本区域,或者在家庭户内使用配额抽样【Quota Sampling】或其他随便选中的单元。

一组广泛的抽样设计都可能符合随机抽样的定义,包括使用多阶抽样【Multistage Sampling】、比例或不成比例分层抽样【Stratified Sampling】、整群抽样【Cluster Sampling】和系统抽样【Systematic Sampling】。

实施随机抽样,还必须作出相当大的努力来极大化调查回答率。若没有这样做,抽样偏差【Sampling Bias】就可能轻易地被无回答偏差代替。有时,这被看成随机抽样的一个弱点,因为每份调查响应的成本通常比非随机方法要高很多(特别在面对面访谈的情况下)。随机抽样法的替代选项是目的抽样法,其中配额抽样法最为人知晓。

——Peter Lynn

(王玥译　赵锋校)

参考文献

Barnett, V. (1974). *Elements of sampling theory*. Sevenoaks, UK: Hodder & Stoughton.

Brewer, K. (2002). *Combined survey sampling inference: Weighing Basu's elephants*. London: Arnold.

European Social Survey Central Co-ordinating Team. (2001). *European Social Survey (ESS): Specification for participating countries*. London: Author.

Eurostat. (1998). *Labour force survey: Methods and definitions* (1998 ed.). Luxembourg: Author.

Moser, C. A., & Kalton, G. (1971). *Survey methods in social investigation* (2nd ed.). Aldershot, UK: Gower.

Park, A., & Jowell, R. (1997). *Consistencies and differences in a cross-national survey*. London: SCPR.

随机变量(Random Variable)

在统计中,一个变量既可以是固定的也可以是随机的。随机变量是一个不能预测其观测值或重复观测时必然得到相同值的变量。一个固定变量则不因为重复测量而改变取值。两类变量的区别可能相当灵活,但根据具体的背景,还是可以清楚地判断一个变量是固定的还是随机的。统计推论(假设检验【Hypothesis Testing】和置信区间【Confidence Interval】)估计仅适用于随机变量。

我们可能认为一把码尺的长度是一个固定变量,因为长度不变。但是,物理学家很早就确认:同一把码尺经多次测量,每次测得的结果不同。因此,长度是一个随机变量,且实际长度(无论其如何得到)同观测长度间的差值可当作测量【Measurement】误差。

在变量关系的研究中,自变量【Independent Variable】常被视为固定变量,而因变量【Dependent Variable】则被视为随机变量。在一项教育程度(以已完成的教育年限来测量)与收入关系的研究中,教育程度被视为固定变量,而收入被视为随机变量。无论测量多少次,某人的教育程度保持不变;受教育16年的人始终取相同的值(除非他们去读研究生)。教育程度相同的人则可能有不同的收入,并且,对每个具有特定受教育水平的群体而言,他们的收入值将散布于某个收入均值两侧。

在一个回归【Regression】分析中,因变量(随机变量)的取值会围绕每一个自变量(固定变量)的取值在回归线上对应的点,呈现高低变化。原因是,我们使得围绕回归线分布的因变量同回归线的垂直均方和【Sum of Squares】的值最小,而非使得散点图上因变量的取值在水平方向或垂线方向上具有最小的均方和。最小化因变量的均方和使我们能够将回归分析的机理应用于这些数据【Data】,以获得相应的结果,如回归线、相关【Correlation】系数和 p 值【p Values】。通行统计软件包中所有普通回归分析的公式,其自变量都是固定变量,因变量则是随机变量。

如果回归分析中的自变量和因变量都被视作随机变量,那么会发生什么呢?因为有时有理由将自变量和因变量都视作随机变量。在这种情况下,所需公式的数学推导就变得更加困难,而且也不能得到与固定的自变量和随机的因变量相似的简单公式。于是,大多数情况下,人们会忽视这个问题,使用适于普通回归的公式,并假定这样做不会对分析和结果产生根本影响。

——Gudmund R. Iversen

(马妍译　赵锋校)

参考文献

Iversen, G., & Gergen, M. (1997). *Statistics: The conceptual approach*. New York: Springer-Verlag.

Moore, D. S. (1997). *Statistics: Concepts and controversies*. New York: W. H. Freeman.

Utts, J. M. (1996). *Seeing through statistics*. Belmont, CA: Duxbury.

随机变差(Random Variation)

变差的研究是统计的核心观念。变差意味着若再次测量某对象,会得到一个不同的结果,而且我们无法预测未来任何一次观测的结果。只要存在随机变量【Random Variable】,根据定义,就存在与之相应的变差。

因此产生了两个问题:一是我们应当怎样测量变差?二是一个变量中的变差有多大比例是由其他变量引起的?测量变差的一个简单方法是计算它的标准差或者其他相关的统计量【Statistic】。其他变量对变差的贡献度,通常用平方和或者多元相关系数【Multiple Correlations】来衡量。

获知一变量变差量级的一种方法是将其作为因变量【Dependent Variable】,并引入一个或多个自变量【Independent Variable】,也就是说,设立一个回归【Regression】分析。这类分析将得到各种平方和。总平方和即因变量方差【Variance】中的分子,测定了因变量的总变差。两个原因导致了因变量取值的不同。一是因变量方差中包含了自变量的不同取值;二是除了自变量以外,还有其他变量对因变量的效应。所有其他变量的联合效应被称为残差【Residual】变量(或误差【Error】变量)的效应。

总变差的部分可归之于自变量。可以使用估得的回归方程算出一组因变量的预测值。那么,为什么这些预测值各不相同?它们之所以不同是因为它们由自变量计算得出,而自变量有不同的观测值。因此,因变量预测值的变差是由于自变量的变化。该变差的量可通过计算回归平方和【Regression Sum of Squares】来衡量。依据由自变量产生的因变量的变差大小,回归平方和给出了自变量对因变量的效应。

总平方和减去回归平方和的差值给出了残差变量的效应,或者说,由所有其他变量产生的总变差的量。这个差值被称为残差平方和【Residual Sum of Squares】,或误差平方和(即使它没有任何错误可言)。残差平方和也可以这样计算:先计算因变量的观测值与预测值间的差值,再将这些差值的平方累加。

每次我们在模型中新增一个带有参数的变量,残差平方和将保持不变或缩小。但是,参数越多,模型就越不俭省(parsimonious)。在极端的情况下,当引入的参数数目等于观测数目时,残差平方和等于零。然而,由于只是以另外的 n 个量(参数)代替了现有的 n 个量(数据点),我们实际并无所获。

——Gudmund R. Iversen

(马妍译　赵锋校)

参考文献

Iversen, G., & Gergen, M. (1997). *Statistics: The conceptual approach*. New York: Springer-Verlag.

Moore, D. S. (1997). *Statistics: Concepts and controversies*. New York: W. H. Freeman.

Utts, J. M. (1996). *Seeing through statistics*. Belmont, CA: Duxbury.

随机游走(Random Walk)

随机游走是一个随机的【Stochastic】过程,有时用于分析时间序列【Time-Series】数据模型。也许,随机游走最著名的应用是连续多日的股票市场价格。

假设函数[u_t]是一个相互独立且同分布的随机变量【Random Variable】序列。于是,[u_t]有固定的均值μ和方差σ^2,并且u_t和u_{t+k}($k = \pm 1, \pm 2, \cdots$)的协方差等于0。由上述属性的随机变量生成的过程被称为一个"纯随机过程"或者白噪声【White Noise】。

根据查特菲尔德(Chatfield,1996)的精确阐述,一组随机变量[y_t]被视作一次随机游走,假设:

$$y_t = y_{t-1} + u_t \tag{1}$$

在实践中,假定时间序列从0开始,即$t=0$。于是,$y_1 = u_1$,$y_t = \sum_{i=1}^{t} u_i$。

相应的,期望值$E(y_t) = t\mu$,方差$\mathrm{Var}(y_t) = t\sigma^2$。由于均值和方差随$t$变化,因此随机游走是非静止的。然而,通过一阶差分,可以得到$(y_t - y_{t-1}) = u_t$,这是一个纯随机过程。

当式(2)中$\alpha = 1$时,随机游走就是一阶自回归【Autoregressive】过程的一个特例。

$$y_t = \alpha y_{t-1} + u_t \tag{2}$$

关于随机游走和整合移动平均自回归模型(Autoregressive Integrated Moving Average, ARIMA)间的联系详见博克斯、詹金斯和赖因泽尔(Box, Jenkins, & Reinsel, 1994)的论述。

——Richard A. Berk

(马妍译　赵锋校)

参考文献

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis: Forecasting and control* (3rd ed.). New York: Prentice Hall.

Chatfield, C. (1996). *The analysis of time series: An introduction*. New York: Chapman & Hall.

随机系数模型(Random-Coefficient Model)

参见混合效应模型【Mixed-Effects Model】、多层次分析【Multilevel Analysis】。

随机效应模型(Random-Effects Model)

方差分析【Analysis of Variance, ANOVA】可用于研究因变量Y如何依赖于一个或多个因子。这里,一个因子被界定为一个类别的自变量,或是一个具有定类测量层次【Level of Measurement】的解释变量。在一个随机效应方差分析模型(也称为第二类方差分析模型)中,每个因子的类别的效应被设定为随机变量,也就是一个带有随机

参数的模型。具体而言,当数据集中一个给定的因子的取值为 1,2,…,K 时,对于个案 i,该因子的取值记作 $k(i)$,相应的,在 Y_i 的线性模型中,该因子在个案 i 上的随机效应定义为 $U_{k(i)}$,同时假定 U_1, …, U_K 是独立同分布的随机变量【Random Variable】。通常,假定这些随机变量服从一个正态分布。如果这个因子的所有类别被视为来自某个总体的随机样本,并且如果统计推论的目的在于获得这个总体的结论,这样的模型就是恰当的。该模型与在其他词条中介绍的固定效应模型【Fixed Effects Models】截然不同。随机效应方差分析模型也被称为方差成分模型【Variance Components Models】。

适于应用随机效应模型的有两类实例:一类是小学生教育成就的研究,其中教师是一个重要的因素,同时,研究者的兴趣主要在于了解教师总体的影响,而非研究数据涵盖的教师的影响;另一类是应用概化理论【Generalizability Theory】框架和评判者信度【Interrater Reliability】的研究,其中对个体的评分得自几个观测者,研究的兴趣之一在于考察观测者(从观测者总体中随机抽选)对测量变异度的影响,而测量变异度则得自所有观测者对某个人的评分的均值。

随机效应同样出现在其他更复杂的模型中,如广义线性模型【Generalized Linear Models】或非线性回归模型中。它们的定义相同:每个因子的各个类别的效应被设定为随机变量。

随机效应模型这一术语仅用于所有因子都有随机效应的模型。同时包含随机效应和固定效应的模型被称为混合效应模型【Mixed-Effect Models】,这类模型比单纯的随机效应模型更常用。上文定义的随机效应可以被视作一个给定因子的各个水平的随机主效应。它可一般化为一个数值型解释变量的随机系数(参见随机系数模型【Random-Coefficient Models】)。对嵌套层次的因子来说,包含随机系数及固定效应的模型也被称为分层线性模型【Hierarchical Linear Models】。这些模型是多层次分析【Multilevel Analysis】的基石。

——Tom A. B. Snijders

(马妍译　赵锋校)

参考文献

Cobb, G. W. (1998). *Introduction to design and analysis of experiments*. New York: Springer-Verlag.

Jackson, S., & Brashers, D. E. (1994). *Random factors in ANOVA* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-098). Thousand Oaks, CA: Sage.

Neter, J., Kutner, M., Nachtsheim, C., & Wasserman, W. W. (1996). *Applied linear regression models*. New York: Irwin.

Scheffé, H. (1959). *The analysis of variance*. New York: Wiley.

随机区组设计(Randomized-Blocks Design)

作为实验【Experiment】的结果,在各个实验单元彼此之间,因变量【Dependent Variable】(反应变量)的值可能是不同的。正是这些值的方差给我们提供了关于自变量【Independent Variable】(处理变量)和或许存在的其他变量的效应的线索。一般可通过用每个观测值减去均值,再对所有的差值取平方,最后加总平方项的方式,测得反应变量的值的方差。这一方法给出了总平方和,$\text{TSS} = \sum (y - \bar{y})^2$。

使用不同的设计,可以以不同的方式建立实验。然而,任一设计的目标都在于,从其他变量中分离出处理【Treatment】变量的效应。我们希望,响应变量的值的方差是由处理变量引起的,而不是其他的无关变量。在随机区组设计中,相似的实验单元被集中在称为区块的各组内。实验组和控制组【Control Groups】都包含区块。

说明和例子

例如,心理学家可能怀疑性别变量会影响知觉实验的结果。不使用区块,实验包括一个实验组和一个控制组。实验后,得到一个实验组的均值和一个控制组的均值。从两组均值之间的差值,以及它们与总均值的差值可以了解处理的影响。测量这种影响的一种方法是计算处理的总平方和。

所有实验单元的总平方和同处理的总平方和间的差值即残差平方和。这一总和给出了所有其他非处理变量的联合影响,其中包括性别的影响。为了确定统计显著【Statistical Significance】,通常的做法是,将处理平方和和残差平方和除以各自的自由度,从而得到各自的均方和。两个均方和可用于计算 F 统计量的值,即 F=处理均方和/残差均方和。F 值越大,p 值【p Value】越小,结果就越显著。

要得到较大 F 值的一个方法是有一个大的分子,而另一种方法是有一个小的分母。我们难以改变分子的大小,但是可能减小分母。因变量的值有所不同的部分原因可能是人的性别不同。如果性别的影响在某种程度上是存在的,那么从残差平方和中取出此一方差来源的方差量,就可能得到一个更小的、新的残差平方和,并由此得到一个更小的 F 分母,从而也将得到一个更大的 F 值。

区块设计的想法即源于此。通过指明每个受试主体的性别,可能安排每类性别的平均知觉得分,以及处理组/控制组的平均知觉得分,见表 1。

表 1 区块和处理的平均得分

区块	处理组	控制组	均值
女性	$\bar{y}_{ft}$	$\bar{y}_{fc}$	$\bar{y}_f$
男性	$\bar{y}_{mt}$	$\bar{y}_{mc}$	$\bar{y}_m$
均值	$\bar{y}_t$	$\bar{y}_c$	$\bar{y}$

对均值的上述安排不会改变处理的平方和的计算。它仍只包含处理的均值和控制组的均值。因此,F 的分子同没有区块化所计算的结果是一致的。然而,现在有可能计算出区块的平方和。该值可根据两个性别组的均值、每一区块观测的数量,以及总均值获得。区块的平方和同处理的平方和非常相似。从下面的公式可得:区块的平方和= $\sum n_f(\bar{y}_f - \bar{y})^2 + \sum n_m(\bar{y}_m - \bar{y})^2$。接下来,用旧的残差平方和减去区块的平方和,得到一个新的、比原来小的残差平方和。再适当地调整各平方和的自由度【Degrees of Freedom】,得到一个新的,比没有区块化时更小的 F 分母。这就得到了一个更大的 F 值和更小的 p 值。

此处,关键在于可以"提纯"残差平方和。残差变量可以被视为所有其他变量的净效应,区块变量是其中的一个,而我们已经可以取出这个变量的效应。因此,现在的残差就由每一个观测值同适当的双下标(处理/控制,女性/男性)均值之间的差值确定;而在没有采用区块设计时,残差等于每一个观测值同处理组的均值或同控制组的均值间的差值。

历史的说明

实验计划背后的很多思考可以追溯到罗纳德·费希尔爵士的关于农业数据的著作中

开拓性的想法。他指出,在将田块分成不同的实验单元时,由于海拔的差异和阳光照射量的差异,田地中某一部分的实验单元可能不同于田地中另一部分的实验单元。如果不以某种方式将这些因素纳入考虑,它们将可能影响不同实验单元的产量,同时影响产量值的方差。解决这个问题的一个方法是通过使用随机区组设计。不是简单地把实验单元分入实验组和控制组,而是首先把相似的土地单元归入不同的组(他称为区块)。

应用

没有区块化,分析只包含一个简单的、单因方差分析;使用区块设计和表 1 中计算的区块变量的均值,分析就是一个没有考虑处理-区块交换效应的二维方差分析。

——Gudmund R. Iversen

(王玥译　赵锋校)

参考文献

Agresti, A., & Finley, B. (1997). *Statistical methods for the social sciences* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Cobb, G. W. (1998). *Introduction to design and analysis of experiments*. New York: Springer-Verlag.

Moore, D. S., & McCabe, G. P. (1998). *Introduction to the practice of statistics*. New York: W. H. Freeman.

随机对照试验(Randomized Control Trial)

随机对照试验(RCT)是一种用于评估的研究设计。RCT 处理的问题是,介入是否会“起作用”,以及在多大程度上会“起作用”。一次试验即一次检验,而术语有控制表明研究设计中包含一组或多组比较组,此外,随机化(randomized)是指用一种随机分配的方法来产生介入组和比较组。随机分配的应用是 RCT 同其他可行的评估设计相区别的唯一特征。RCT 的作用在于,就介入的影响,减少被误导的机会——即由于用其他方式选取介入组和控制组而带来的未被测量的因素所产生的误导。

RCT 何以有用

对研究者、科学家、政策制订者和一般大众而言,评估不同类型介入的有效性是一项主要的挑战。对有行为问题的儿童的密集社会工作介入是否实际帮助了这些儿童和他们的家庭?计算机辅助学习对于儿童读写能力有什么样的影响?把犯人介绍给受害者能否确保更低的再犯率?乳腺癌和前列腺癌的筛查项目是否减少了这些疾病的死亡率?回答这些问题需要使用一种有效的研究设计,而这种设计能够使得问题中的介入影响尽可能地同其他影响——包括介入接受者的社会特征,以及一定时期中和一定社会背景下持续存在的影响要素——分解开来。RCT 可以为政策和实践决策,就各种介入的相对有效性,提供严格的证据。然而,如同在任何研究中一样,一项好的研究设计,仅当它陈述了一个重要的问题,且有很好的操作时,才会有最佳的效果。

RCT 的历史

控制估计介入效应中存在的偏差的努力至少可以回溯到 3 个世纪前,同时发生于医学和社会科学领域,而当时努力的基础在于,人们认识到意图良好的介入却可能带来伤害。从 17 世纪开始,医生建议抽签决定

哪个病人将接受某种特定治疗。而进行同类相比这一需要在18世纪获得了广泛的认可,比如,海军外科医生詹姆斯·林德针对坏血病的著名研究,在这个研究中,具有相同状况的病人被分别施以不同的治疗手段,而那些拿到橙子和柠檬的病人恢复得最快。1920年后不久,运用交替或"随机"分配产生比较组的方式变得更加普遍,而对随机分配过程的第一个详细描述则出现于1930年代。术语"控制"一词自1870年代开始用于实验心理学,以指称进行比较的一个标准,即为检验得自某项实验的推论而实施的比较的标准;该词来自"反向-登记",指用以核查官方账目的重复簿录。

尽管当前RCT主要与健康照料领域有关,社会科学家也同样考虑过,设计介入评估,从而能够可靠地回答影响效果的问题。1920年代、1930年代以及1940年代中,研究者,特别在北美,教育、社会福利和公共政策领域内的研究者,广泛地使用了实验研究设计,同时,实验社会学成了一个重要的研究领域。随后是1960年代到1980年代的"评估的黄金时代",在此期间,美国的评估者——通常有政府背景,开始运用RCT来评估范围宽广的公共政策,其中包括收入维系和住房津贴方案、弱势工人支持项目,以及对刑满释放人员的介入。

RCT和其他评估有效性的研究设计

一项RCT可以通过比较两个群体的确定结果(变量)来检验某种介入的效应,其中,一个群体被施以一种介入,而另一个群体则没被给予。设计中,使用某种随机分配(机会)的方法,如抛掷硬币或利用一种计算机随机化程序,将一项介入的潜在接受者分为介入组或控制组。无论个体,还是群体(如学校、医院或社区)都可以成为随机分配的对象("整群"随机化试验)。整群RCT在社会科学领域可能特别有用。

经验上,RCT的结论通常不同于那些非随机化或非控制研究的结论(例如,常见的只对一个组做前测和后测的设计)。RCT通常比其他类型的研究产生更低的有效性估计。在社会科学和政策中,还有很多例子表明RCT的结论与观察研究的结论不同。如针对幼儿日间照顾的一项RCT,即佩里幼儿园项目(the Perry Preschool Project),通过27年的追踪研究,表明这一介入在教育成就、未计划怀孕、犯罪行为等被测变量上均有良好的长期效应,而与之相反的是,许多观测研究的发现指日间照顾具有有害效应。

目前关于RCT的论争

关于RCT方法论的最近工作是由那些联系在Cochrane协作网(the Cochrane Collaboration,以流行病学家阿奇·科克伦(Archie Cochrane)命名的协作网络,他本人强烈地主张基于RCT的研究证据的综合)之下的研究者间开展的。一个同类网络,Cambell协作网(the Campbell Collaboration),建立于2000年,使用同样的原则来分析社会介入的证据。该网络得名于唐纳德·坎贝尔,他是一名方法论专家,并且在1966年与斯坦利的著作中,建立了在社会科学中应用RCT的原则和程序。

遗憾的是,由于"RCT"一词同医学的联系,这就导致一些社会科学家放弃了这一方法。然而,对RCT的伦理的异议忽略了不能适当回答研究问题的研究设计的不道德性。这些异议同样也疏于承认无控制实验的常规,即这些常规使得人们可能受到带有未知的,也许有害的效应的专业或其他介入的伤害。成千上万次已经实施的社会介入的RCT显示,即使在复杂的社会环境中,这一方法也是有效的。对于解决

当前某些最紧迫的社会问题而言,一项开展良好的 RCT 是用以评估相关策略的一种强有力的方法。

——Ann Oakley

(王玥译 赵锋校)

参考文献

Boruch, R. F. (1997). *Randomized experiments for planning and evaluation: A practical guide*. Thousand Oaks, CA: Sage.

Campbell, D. T. (1988). *Methodology and epistemology for social science*: Selected papers (E. S. Overman, Ed.). Chicago: University of Chicago Press.

Oakley, A. (2000). *Experiments in knowing: Gender and method in the social sciences*. Cambridge, UK: Polity.

Orr, L. L. (1999). *Social experiments: Evaluating public programs with experimental methods*. Thousand Oaks, CA: Sage.

随机化应答(Randomized Response)

在一个调查【Survey】情形中,当问及敏感的或有威胁的问题时,尽管有保密的保证,回答人还是会经常说谎。实际上,对大多数人而言,推进社会科学研究的理由并不足以促使他们透露那些令其尴尬,甚或可令其成为罪犯的信息。

随机化应答是一种旨在保护回答人而将某种隐瞒结合到调查工具中的数据收集策略。这就是根据某些随机过程,鼓励回答人给出不明确的答案。好像流行的游戏节目危险边缘(Neopardy),访谈员知道回答人的答案,但是不确知与之关联的问题。这一方法最简单的形式(Warner, 1965),是根据某种随机设置(randomizing device)的结果,比如掷骰子,请回答人回答一个敏感的问题或它的反面问题。例如:

掷出 1~2:在过去的一个月中,你用过毒品吗?

掷出 3~6:在过去的一个月中,你戒除了毒品吗?

要点在于,只有回答人自己知道掷骰子的结果,因此,回答"是"或"否"既可以是对毒品使用的承认,也可以是否认,而只有回答人才知道确切的答案。不过,在总计中,毒品使用的普及率可以从作出肯定回答的样本【Sample】百分比中估计出来(其中一部分人承认使用毒品,剩余部分声称戒除毒品)。

肯定回答的比例(π)是 $\lambda = p\pi + (1-p)(1-\pi)$,其中,$p$ 是上述随机设置的二项【Binary】概率(如出现 1 个或 2 个点的概率为$p=1/3$,λ 是所求的普及率参数,例如,过去一个月中使用毒品的百分比)。然后,可根据肯定回答的样本比例,对考查的普及率的敏感度作出估计,即

$$\hat{\pi} = \frac{\hat{\lambda} + p - 1}{2p - 1}$$

尽管其简单性使其显得灵巧,但是,将一个敏感问题与它的反面问题配成一对可能引起回答人对于这种统计策略的怀疑。一个更有效的,被称作无关问题法的策略(Horvitz, Shah, & Simmons, 1967),则通过包含一个无威胁的备选问题,来增强回答人的心理安慰。回答人仍然依据某种随机设置(如投硬币)的结果,被要求回答一个敏感性的问题或另一备择问题。例如:

头面:过去一个月中,你用过毒品吗?

背面:你的社会保险号码最后一位是奇数吗?

在总计中,若已知半数回答人会被引导回答社会保险号码的问题,同时,其中一半人会回答“是”,我们可以分别计算出相应的回答分布。为了估计毒品使用的普及率,要从观测到的回答分布中剔除对非敏感问题作出肯定回答的期望数。

自沃纳的开创性论文发表以来,许多使用多个问题、多组别及不同随机设置的备选方法已经不断被提出来(进一步讨论参见:Fox & Tracy, 1986; Chadhuri & Mukerjee, 1987)。当然,诸方法的目标都在于,设计出一个能让回答人对敏感部分产生免疫的调查问题,这样的设计既要令回答人信服,还要兼具统计上的效率。为了减少规避回答的偏差,由随机设置和多问题使用产生的不确定可能在统计精度上引起一定程度的损失。

随机化应答法也已经扩展到数量型答案上。如,请回答人先从一个罐子中挑出一个有颜色的球,再根据球的颜色,请他回答下述两个问题中的一个:

蓝球:过去一个星期中,有几天你用过毒品?

红球:你的社会保险号码的最后一位是?

观测到的回答的均值(Z)是对敏感问题(X)回答的频次均值和对替代问题(Y)回答的数字均值的加权平均,其权数是与随机设置相关的概率(p)。由此可以很容易得出敏感问题的估计均值,即

$$\hat{\mu}_X = \frac{\overline{Z} - (1-p)\mu_Y}{p}$$

根据非敏感特征分布的知识,以及随机设置的知识,人们可以很容易从完整的答案汇集中得出敏感问题的均值和百分比分布。

在所有设计中,只有随机化应答策略允许在无须暴露个体回答人的任何确切特征的前提下,对总和参数作出估计。即便没有个体水平的得分,通常施用于个体得分的各种多变量【Multivariate】技术实际上仍然有效。而通过将随机应答数据作为具有已知(或可估)分布属性的测量误差【Measurement Error】的一个特例,相关【Correlation】和其他以关联【Association】为基础的统计量都是可以估出的。

自这一技术引入以来,随机化应答已经成功地用于许多社会科学领域,实现了对诸多敏感行为的估计,如学术欺诈、堕胎、组织内不当行为及举报、不良商业行为、性侵犯、儿童性骚扰犯罪、有艾滋病风险的性行为、药物滥用,以及其他形式的犯罪活动(Daniel, 1993)。

尽管有其统计上的吸引力,但是随机应答技术并没有真正地为应用研究共同体热情接受。虽然这一技术在数学上看起来是合理的,有研究者却对回答人在这一技术实施过程中的理解和信任提出了疑虑。不过,几个不同领域的确证已经给出了肯定的结果。许多研究检验了随机化应答技术相对于已知数据的表现,还对比了由随机化应答得到的估计值和其他相关方法得到的结果,如直接提问(Armacost, Hosseini, Morris & Rehbein, 1991; Stem & Lamb, 1981; Tracy & Fox, 1981; Van der Heijden, van Gils, Bouts, & Hox, 2000)。随机化应答技术始终表现得优于传统访谈技术,即针对社会反对的或不法行为的发生率,得出较高的估计值。

虽然在减少回答偏倚(倾向于给出不真实的回答)方面,随机化应答技术已经表明是有效的,但是就这种技术是否能帮助减少无回答偏倚【Nonresponse Bias】(拒绝参与回答),仅有混合的证据。当没有人知道他们正在回答哪一个问题时,回答人可能会更

加放心地去诚实回答问题,不过,认为这个技术太复杂或太费力,这也会成为拒绝参与的理由。在最终的分析中,随机化应答技术也许最适用于相对有教养的人群——这些人群中的回答人能够理解和欣赏这一技术给予他们的保护。即使如此,有些回答人还是会像担心他们自己的隐私一样,担心他们的回答(即以总计的形式)对他们所在的组织或职业产生影响。

无论何时要成功地应用随机化应答法,都有一些需要注意的重要考量。第一,为了平衡更大的不确定性(即抽样误差【Sampling Error】)——来源于该技术本身带来的随机误差,样本量就必须相对大些。第二,由于增加的误差,不建议运用这一技术来考察罕见事件,特别是那些普及率不到10%的事件。尽管有如上顾虑,还是存在一些社会科学质询,由于它们处理的问题是如此棘手,以至于一种像随机化应答这样的方法是完全不可或缺的。

——James Alan Fox
Maria Tcherni
(王玥译　赵锋校)

参考文献

Armacost, R. L., Hosseini, J. C., Morris, S. A., & Rehbein, K. A. (1991). An empirical comparison of direct questioning, scenario, and randomized response methods for obtaining sensitive business information. *Decision Sciences*, *22*, 1073-1090.

Chadhuri, A., & Mukerjee, R. (1987). *Randomized response: Theory and techniques*. New York: Marcel Dekker.

Daniel, W. W. (1993). *Collecting sensitive data by randomized response: An annotated bibliography* (2nd ed.). Atlanta: Georgia State University Business Press.

Fox, J. A., & Tracy, P. E. (1986). *Randomized response: A method for sensitive surveys*. Beverly Hills, CA: Sage.

Horvitz, D. G., Shah, B. U., & Simmons, W. R. (1967). The unrelated question randomized response model. In E. D. Goldfield (Ed.), *Proceedings of the Social Statistics Section* (pp. 65-72). Washington, DC: American Statistical Association.

Stem, D. E., Jr., & Lamb, C. W., Jr. (1981). The marble-drop technique: A procedure for gathering sensitive information. *Decision Sciences*, *12*, 702-707.

Tracy, P. E., & Fox, J. A. (1981). The validity of sensitive measurements: A comparison of two measurement strategies. *American Sociological Review*, *46*, 187-200.

Van der Heijden, P. G. M., van Gils, G., Bouts, J., & Hox, J. J. (2000). A comparison of randomized response, computer assisted self-interview, and face-to-face direct questioning: Eliciting sensitive information in the context of welfare and unemployment benefit. *Sociological Methods & Research*, *28*, 505-537.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, *60*, 63-69.

随机性(Randomness)

在统计学里,随机性是一个基本概念,而在此基础上,统计推论【Statistical Inference】得以建立。利用适当的随机性,可以进行假设检验【Hypothesis Testing】,并为未知参数设定置信区间【Confidence Intervals】。

当一次试验或一次观测的结果在实施前,不可能被预测到时,随机性即存在。

这一概念本身存在的范围要超过它在统计学中的应用。许多简单的例子都有随机性的存在,如赌博、掷硬币或骰子,以及玩纸牌。在掷硬币之前,我们不能准确地说出,出现的是头面还是背面。这就使得"掷硬币"成了一个有两个值的随机变量【Random Variable】,除非我们特别精于掷硬币这种游戏。统计分析中的大多数因变量都是随机变量。

随机变量的反面是其值为固定值的变量。大多数自变量【Independent Variable】被认为是有固定值的变量。我们把教育(比如说受教育年份)看作是固定的。同时,收入被认为是一个随机变量。只有通过随机抽样,才能获得具有固定受教育年限的人群,进而研究各受教育年群组的收入分布【Distribution】。

说明

在统计学里,有两件事情需要随机性,即数据【Data】收集和统计推论的应用。对观察数据而言,一个样本【Sample】中的诸单元(居民、市民等)都必须从更大的总体【Population】中随机地选取。选取样本可以使用不同的抽样方案,例如简单随机抽样【Simple Random Sampling】、分层抽样【Stratified Sampling】、整群抽样【Cluster Sampling】。只有在使用随机抽样的情况下,才能运用统计公式计算 p 值【p Values】或置信区间。因为,这些公式的数学推导所根据的假设是,数据是根据适当的统计抽样程序而汇集的。实际上,流行的统计计算包中的大部分公式仅在简单随机抽样的假设下适用。

实验【Experiment】数据也需要随机性。受试单元必须以随机的方式被分入实验组和控制组。多种实验设计均能保证所得数据具有适当的随机性。

历史说明

早在《旧约全书》和《维京传奇》中就有提到随机性。书中提到运用某些随机机制来作决定。对随机性更集中的研究始于17世纪,那时有些绅士赌徒想要为用纸牌和骰子作工具的机会游戏建立相应的投注规则。这一企图催生了致力于发展相应数学工具——我们今天称之为概率【Probability】理论——的数学家们。

随机性在科学哲学中不但有其基础,还引发出若干议题。例如,有理由认为,存在着随机性程度的差别。比如,对某个现象知道得越多,它就越少具有随机性。在贝叶斯统计中,这点引出了具有减少不确定范围的后验分布。

同时,一个真正擅长掷硬币的人可能以这样一种方式投掷硬币,即可以对他每次投掷的结果,出现头面还是背面,作出准确预测。相似地,在多元量【Multivariate】统计分析中,我们可以在一个回归【Regression】分析中增加变量,从而减少剩余变量的效应——正是剩余变量被认为代表着因变量的随机性。如果可以增加变量,致使多元相关【Multiple Correlation】系数 R 等于 1,那么就不存在随机性了。

应用

在纯真空的世界中,一个物体落下一定距离所用的时间就不存在随机性,而物理学也已经有了相关变量的方程。若不存在随机性,变量间的关联就被表示成一种自然规

律的样式。此类规律常令那些试图以相似方式建立变量间关系的社会科学家称羡，不过，他们的努力至今也没有成功。

随机性赋予概率理论以基础。当我们不知道一项试验或一次观测的结果时，可以尝试把不同的结果同各自的概率连在一起。实际上可以说，概率理论构成了对随机性的研究。反过来，概率理论又为统计推论提供了基础。

我们如何才能造出随机性，并由此收集随机数据？这点说起来容易，做起来却很难。样本必须用以随机数为基础的方法来抽取，比如，简单随机抽样，或是多阶分层抽样。一个极端缺少随机性的例子是美国越战期间的第一轮征兵抽签，其中对生日的选取不是随机进行的，而是出生年靠后的年轻人比那些出生年较早的人更有可能被抽中。

例子

我们怎样知道随机数真的是随机的？我们说一组数字是一个随机数集合是什么意思？如果随机数仅用于抽取随机样本，上述疑问就不是一个非常重要的问题。但是，随机数也会用于其他背景中，其中最重要的是创制秘密代码的密码工作。同样，国家彩票也利用随机性，这时，任何号码，只要它不是随机中奖号码，都不会为公众接受。

这意味着，计算机科学家和其他科学家在试图确定有一个真正的随机数集的真正意义的同时，已经就如何造出随机数集这一问题进行了许多思考。一个真正的随机数生成程序应当如此，即如果用此程序第二次生成一个随机数序列，那么我们应当得到一个与第一次生成的随机数序列完全不同的数列。但是，大多数计算机程序有赖于所谓的种子数(seed number)来生成随机数。如果将同一种子数用于第二次生成过程，我们就会得到同样的数字集合，也就意味着这些数并不随机。同时，在数列中，0~9 中的任意一个数应当同其他数出现得一样频繁。但是，即便 0~9 的所有数序都是等可能的，我们还是可能得到一个，比如说，全都是 2 的序列。即使这个序列是随机产生的，我们还是会犹豫最好不要使用这样一个序列。

人类并不擅长预测未来，即使只是几分钟后的未来。这意味着，我们发现我们生活在一个随机的世界中。为了理解这种随机性，需要在随机性中发现规律性，进而创造出某种程度的秩序。

——Gudmund R. Iversen

（王玥译　赵锋校）

参考文献

Beltrami, E. J. (1999). *What is random? Chance and order in mathematics and life*. New York: Copernicus.

Bennet, D. J. (1998). *Randomness*. Cambridge, MA: Harvard University Press.

Keren, G., & Lewis, C. (Eds.). (1993). *A handbook for data analysis in the behavioral sciences: Methodological issues*. Hillsdale, NJ: Lawrence Erlbaum.

全距(Range)

全距是一个变量【Variable】最大值和最小值之间的距离。例如，设变量为年龄，研究

中最年轻的是26岁,最年长的是93岁,那么年龄变量的全距为26~93,或表示为67(93−26=67)。全距是一个变量离散【Dispersion】度的最方便的量度。然而,它却并非离散度的最终量度,因为它极易受到游离值【Outliers】的影响。在例中,设若低于最高年龄93岁的次高年龄是61岁,那么年龄93的出现增大了变量的全距,这一数据所暗示的高龄人口数要多于数据实际给出的高龄人口数。

——Michael S. Lewis-Beck

(马妍译　赵锋校)

等级排序(Rank Order)

等级排序是指一组分数的属性,这组分数可以从最低排到最高,或相反,从最高排到最低。一组分数15, 14, 8, 11, 12的等级排序可以按照从最低到最高,排列为8, 11, 12, 14, 15,或者按照从最高到最低,排列为15, 14, 12, 11, 8。从形式上说,若将一组分数按照从最低到最高的方式予以等级排序,那么对任意给定分数 X_j 总有 $X_{j-1} \leq X_j \leq X_{j+1}$;若按照从最高到最低等级的方式,对于任意给定分数 X_j,总有 $X_{j-1} \geq X_j \geq X_{j+1}$。

有时,依据一组分数的等级排序,可以为每个分数分配一个秩(rank)。秩是连续的有序数码(通常是整数)。这些数码反映了一个分数在一组依等级排序的分数上的相对位置。例如,可以赋予分数8, 11, 12, 14和15如下的秩:

表1

得分	秩
15	5
14	4
12	3
11	2
8	1

若最低的分数被给予秩1,秩3就意指,该分数相对于最低分数是第三高的分数。秩4意指,该分数相对于最低分数是第四高的。有时,也可以赋予最高分数的秩为1,最低分数的秩为 N,其中 N 是所有排序分数的总数目。

一组分数中,相同的分数可以出现两次,如8, 11, 15, 14, 8, 12。仍然可将该组分数按照从最低到最高,排序为8, 8, 11, 12, 14, 15;或者,按照从最高到最低,排序成15, 14, 12, 11, 8, 8。在给重复得分赋秩时,传统的做法是以重复分数所得秩的均值为秩值,见表2。

表2

得分	秩
15	6
14	5
12	4
11	3
8	1.5
8	1.5

其中,赋予两个相同的分数8的秩为 $(1+2)/2=1.5$。

——James J. Jaccard

(马妍译　赵锋校)

融洽关系(Rapport)

融洽关系是指这样一种访谈情形,即访谈员和受访者都感觉舒适,且相互之间协调一致。一种融洽关系能够促使每一方了解对方正在叙说的内容。

——Tim Futing Liao

(王玥译　赵锋校)

Rasch 模型(Rasch Model)

Rasch 模型是题项反应模型族中一个专门且结构简单的模型。对一个题项的二分【Binary】响应[即正确(得分 1),不正确(得分 0)]依赖于一个人的潜特征值和题项的难度。联系得分 1 和潜特征值间的函数是 Logistic 函数,对每个题项有唯一的位置(难度)参数,但有相同的区别效力。

——Klaas Sijtsma

(马妍译　赵锋校)

*也可参见项目反应理论【Item Response Theory】。

率的标准化(Rate Standardization)

自社会科学家开始注视经验世界以来,他们就多少明确地进行比率的比较。但直到 19 世纪末和 20 世纪初,学者才真正开始尝试将可观测到的比率间的差异归之于一些混杂【Confounding】因素的分布差异。例如,我们可能想比较两个地区的犯罪率。如果这两个地区的年龄分布【Distributions】有显著差异,且如果已知问题中的犯罪率随年龄变化,那么两个地区的犯罪率就必然随两个地区年龄分布差异的程度而变化。此即,所有可观测到的犯罪率的差异,其中有一部分必然可归之于年龄分布的差异。而其中这部分的犯罪率差异不能归因于地区间促进或抑制犯罪的固有属性。当然,在为相关的社会政策提供决策信息时,这也会产生明显不同的后果。由此,在比较比率时,一个重要的问题是:可观测到的比率差异在多大程度上可归之于被比较的总体间的所谓组分差异(compositional differences)?

在 20 世纪前半叶,部分解决此难题的方法是考虑,如果被比较的总体具有相同的组分[可以任意确定,但通常根据年龄、种族和(或)性别的分布],那么两个或多个比率间的差异会是什么。为此,研究者经常将某一总体作为组分分布的标准,而这个标准既可以是被比较总体之一,也可以是比较研究以外的其他总体。

然后,剩余比较总体的比率将根据剩余总体和标准总体间的组分差异,予以调整和控制。被调整后的比率就被称为标准化的比率(standardized rates)。标准化的比率,用当前的语言形容是对数线性的。此即,它们不是实际能观测到的,而是被构造出来的,好像可以观测发现它们具有相同的,即所谓标准的组分分布。因此,标准化的比率差异不能再归之于被比较总体间在组分分布上的差异。

20 世纪早期,研究人员通过比较标准化(反事实的)比率与非标准化(观测的或

粗的)比率的差异来计算观测到的差异可归之于组分差异的比例(Groves & Ogburn, 1928)。然而,直到1940年代后期和1950年代早期,此类应用都还是非正式的,直至伊夫林·基塔嘎娃和她在芝加哥大学的同事将这一比较规范化。这项由基塔嘎娃在1955年出版,并被称作“两个比率间差异组成”的技术,成为一项重要的技术突破,精确地表明观测到的未处理的比率可以怎样被表示为,组分差异和特别组分的观测比率的函数。

自伊夫林·基塔嘎娃(Kitagawa, 1955)的奠基性成果以来,许多学者为我们理解比率的标准化和比较作出了有益的方法改进。最重要的一项是克洛格(Clogg, 1978)运用对数线性模型【Log-Linear Models】发展的一般方法,即他称为混杂影响的净化率(purging rates)。克洛格的净化方法——会同克洛格和伊莱亚森(Clogg & Eliason, 1988),廖福挺(Liao, 1989),谢宇(Xie, 1989),克洛格、肖基和伊莱亚森(Clogg, Shockey, & Eliason, 1990)的改进和扩展——提供了一个有效的模型方法,以统计推论工具为基础,以便比较观测到的、标准化的或更一般的净化率,以及计算这些比率的函数关系。

——Scott R. Eliason

(马妍译 赵锋校)

参考文献

Clogg, C. C. (1978). Adjustment of rates using multiplicative models. *Demography*, *15*, 523-539.

Clogg, C. C., & Eliason, S. R. (1988). A flexible procedure for adjusting rates and proportions, including statistical methods for group comparisons. *American Sociological Review*, *53*, 267-283.

Clogg, C. C., Shockey, J. W., & Eliason, S. R. (1990). A general statistical framework for adjustment of rates. *Sociological Methods & Research*, *19*, 156-195.

Groves, E. R., & Ogburn, W. F. (1928). *American marriage and family relationships*. New York: Holt.

Kitagawa, E. (1955). Components of a difference between two rates. *Journal of the American Statistical Association*, *50*, 1168-1194.

Liao, T. F. (1989). A flexible approach for the decomposition of rate differences. *Demography*, *26*, 717-726.

Xie, Y. (1989). An alternative purging method: Controlling the composition-dependent interaction in an analysis of rates. *Demography*, *26*, 711-716.

定比尺度(Ratio Scale)

史蒂文斯(Stevens, 1946)依最低至最高精确度,确定了四个测量层次:定类、定序、定距和定比。定类、定序和定距尺度在本书的其他部分有介绍。

定比尺度与定距尺度有几项共同特性。它们都能区别数量的大小。任意两个相邻数字间的间距所代表的量都是相等的。

定比尺度的独有特征是存在一个绝对的原点或零点。零点能使研究者明确一个量比另一个量大多少或者小多少。定距尺度则无法得此论断[也有人主张定距和定比尺度间的区分无关紧要(Wright, 1997)]。

例如,我的体重是180磅,我妻子重120磅。在比较我们两个人的体重时,可以说,我较重,即一个定序层次的区别。或者,可以说,我比妻子重60磅,一个体重的定距测

量结果。或者，我可以指出，体重是一个定比变量，进而断言，我是妻子体重的 1.5 倍（我是 90 磅重的儿子体重的 2 倍）。此外，如果我们将今天 30 ℉的寒冷同昨天 60 ℉的温暖相比较（使用华氏度量法），可以断言今天低于昨天 30 ℉，但我们肯定不能说，今天只有昨天一半的温暖。

区分测量层次更精细的方法是判别每一种测量层次隐含的不变性。定序尺度可以进行单调【Monotonic】变换，而保留原尺度的信息。定距尺度可以进行线性变换【Linearly Transformed】。定比尺度只能通过乘以常量【Constant】加以变换。此即，只有乘以一个常数项（$X' = aX$）时，它们才能保存原有信息。仅在此种情形下，比值相同。若我增重 180 磅（体重加倍），我妻子增重 120 磅（原来体重乘以 2），我们俩体重的比值保持一致（1.5）。

传统的观点接受了史蒂文斯的论断，认为在每一个测量层次上，都有与之相应的特定的统计方法。例如，适用于定类数据的集中趋势【Central Tendency】的唯一度量是众数【Mode】，中位数【Median】可用于描述定序数据，此外，若数据以定距尺度测得，则用算数均值【Mean】。

定比尺度的数据似乎有内在的价值，但在很多情形中，社会科学家并不追求如此精确。探索性研究通常关注定类和定序测量层次的现象。大多数相关【Correlational】/成因研究寻求定距层次的数据。对定比数据最常见的断言来自对模型交互项【Interaction】的检验。一个自有的原点赋予这些交互项以固定的解释，而以任意设定的零点为中心点的定距数据则难以解释。

尽管，对史蒂文斯的测量层次的分类有表面上的共同接受——至少作为本科生的教学工具——但也有大量的不同观点，认为测量层次不应成为放弃使用特定统计工具的理由（Baker，Hardyck，& Petrinovich，1966；Borgatta & Bohrnstedt，1980；Gaito，1980；Mitchell，1986；Townsend & Ashby，1984；Velleman & Wilkinson，1993）。

一个适当的理由是，测量层次并不是被测对象的固有特征。此即，一个对象的同一测量可由史蒂文斯划分的任何一个测量层次表示出来。同样地，一旦我们承认任何测量中都存在误差【Error】——系统的【Systematic】或随机的【Random】——的可能性就会发现，很难严格地断言某一特定的测量层次就是某对象的确定特征。在确定所研究现象的测量层次时，测量工具也会起作用。

测量的史蒂文斯分类，尽管最常见，但并非唯一的分类框架。例如，莫斯特勒和图基（Mosteller & Tukey，1977），以及克里斯曼（Chrisman，1997）对史蒂文斯的分类提出了许多增补和修正。替代方案的重要性并不在于它们一定就优于或劣于原来的分类，而在于它们清楚地说明，史蒂文斯的分类并非是完全的。

——John P. McIver

（马妍译　赵锋校）

参考文献

Baker，B. O.，Hardyck，C. D.，& Petrinovich，L. F.（1966）. Weak measurement vs. strong statistics：An empirical critique of S. S. Stevens' proscriptions on statistics. *Educational and Psychological Measurement*，*26*，291-309.

Borgatta，E. F.，& Bohrnstedt，G. W.（1980）. Level of measurement—Once over again. *Sociological Methods & Research*，*9*，147-160.

Chrisman，N.（1997）. *Exploring geographic information systems*. New York：Wiley.

Gaito，J.（1980）. Measurement scales and statistics：Resurgence of an old misconception. *Psychological Bulletin*，*87*，564-567.

Koch，W.，Schulz，E. M.，Wright，R.，Smith，R. M.，Lang，S.（1996）. What is a ratio scale? *Rasch*

Measurement Transactions, *9*, 457.
Mitchell, J. (1986). Measurement scales and statistics: A clash of paradigms. *Psychological Bulletin*, *100*, 398-407.
Mosteller, F., & Tukey, J. W. (1977). *Data analysis and regression*. Reading, MA: Addison-Wesley.
Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*, 677-680.
Stevens, S. S. (1951). *Handbook of experimental psychology*. New York: Wiley.
Townsend, J. T., & Ashby, F. G. (1984). Measurement scales and statistics: The misconception misconceived. *Psychological Bulletin*, *96*, 394-401.
Velleman, P. F., & Wilkinson, L. (1993). Nominal, ordinal, interval and ratio typologies are misleading. *American Statistician*, *47*, 65-72.
Wright, B. D. (1997). S. S. Stevens revisited. *Rasch Measurement Transactions*, *11*, 552-553.

理性主义(Rationalism)

这一术语既指一种普遍的哲学立场,也指一组人类行动和社会关系的理论。在西方哲学中,理性主义与经验主义【Empiricism】之间的长期对立在于,我们的知识源自理性还是经验,抑或两者中谁居首。在现代科学哲学中,理性主义者宣称,科学的规律必然是真的,同时,这种必然性的观念有时让他们更接近于实在论和对因果关系【Causality】的实在论说明。然而,理性主义更常用来指称此一原则,即大多数的人为了好的理由而去做,或者可以假定为了好的理由而去做;同时,社会解释应当依据(通常是个体的)理性行动。这就在经济学和其他社会科学领域产生了一项强有力的研究纲领:如果我们认为,卖东西而求最高价是理性的,可以预言,在其他情况都相同时,人们都将这样做。正如哲学家马丁·霍利斯(Hollis,1977)所说,“理性行动是它自己的解释”(p. 21):仅当人们没能理性地去做时,我们才需要一个关于他们的举动的专门解释。当然,存在着理性解释的不同层面:以最低的价格买香烟是理性的,将它们抽完则是非理性的。理性行动和理性选择并不必然是自私的;人们可能因为火车对环境破坏更少,而理性地选择乘火车而非汽车或飞机去旅行。

关于理性行动和理性选择的解释的有效性存在着相当大的争议,尽管它们在政治科学和国际关系领域已经变得越来越突出,甚至在社会学和社会人类学也能发现它们。在作选择时,真的如模型所说的那样计算吗?比如,选择婚姻伴侣、政治政党或宗教信仰?理性行动理论似乎忽略了人类文化和意识形态的丰富特征。另一方面,他们又为社会科学提供了一些有用的工具和解释范式,诸如“搭便车(free rider)”,囚徒困境【Prisoner's Dilemma】,以及来自博弈论【Game Theory】的模型。一般来说,理解【Verstehen】是社会质询中的一项核心元素,即对人们理解自身处境的方式的理解和对人们何以如此这般行事的理由的理解,即便它最终由更多结构化的解释来补充或排除。要记住的是,当我们被认为做错某事时,我们最常被要求去“自我辩解”。

——William Outhwaite

(王玥译　赵锋校)

参考文献

Elster, J. (Ed.). (1986). *Rational choice*. Oxford, UK: Basil Blackwell.
Elster, J. (1989). *Nuts and bolts for the social sciences*. Cambridge, UK: Cambridge University

Press.
Hollis, M. (1977). *Models of man*. Cambridge, UK: Cambridge University Press.
Hollis, M., & Smith, S. (1990). *Explaining and understanding international relations*. New York: Oxford University Press.

原始数据(Raw Data)

我们常哀叹道,"要做的太多,时间则太少"。相同的抱怨也可用于数据【Data】。我们经常发现自己面对着"如此多(原始)的数据,以及这么少(有用)的信息"的状况。对于大多数收集和分析数据的研究者而言,挑战在于从原始材料中萃取出有用的信息。

让我们从一系列描述原始数据的必要区分开始吧。

第一,"数据"是一个复数名词。数据即许多个单个数据。

第二,所有的数据汇集都是不完整的。因为各种各样的限制,总有数据未予收集。时间、金钱,以及两者之合——效率,是我们难以收集能够回答所有问题的可用原始数据的三个常见原因。然而,数据收集逐渐变得更加容易、更加有效率,因此,越来越多的数据被收集。由于缺乏处理这些额外的原始数据的能力,更多的数据并不能增加知识,不过得到提高的计算能力已赋予人们更有效地处理更多数据的能力。

第三,"原始数据"这一表述即含有数据处理水平的意思,即人们不能期望,在没有进一步努力的情况下,当前的数据就能提供有用信息。"原始"暗示着数据还没有以特定的方式得到概括或分析,从而能根据它们被收集的目的释出相应的信息。

第四,原始数据可能是以另一种呈现格式被汇集起来,而这种格式会影响研究者思考该如何处理它们。也许考虑这个问题的最简单方式是,想一想用在计算机处理信息中的各种符号样式。在计算机中,数据常以二进制、十六进制或十进制数字系统的方式呈现出来,或者呈现作(ASCII)文本。数据的组织也同样会有影响:信息可能是以位(bits)、四位组(nibbles)、字节(bytes)、字(words)等组织起来的。

一个人处理过的数据可能是另一个人的原始数据。收集数据的方法和收集到的数据的聚合水平,在分析者决定要如何分析作为原始数据而被聚集起来的材料时,起着重要的作用。

例如,我们可以考虑一下,有一个调查【Survey】问题的答案汇集,而该调查共有5个问题构成一组,目的在于测量公众对堕胎的态度。这些数据最分散的样式是对5个问题中每个问题的个人回答——常见的形式是代表各个答案的数字。一名调查研究者可根据调查问题,利用这些数据建立起个体回答的模式或一致性。这种努力将引出第二个水平上的数据聚合,即造出一个量表【Scale】,以表现对这组问题的回答综合,亦即给每个个体回答人赋予一个回答分数。就所有被访个体的子样本或各个群组,通过总计他们的个体量表分,就可以达到第三个水平上的数据聚合。这些数据聚合的第四个水平可能要报告佩奇和夏皮罗(Page & Shapiro, 1992)所说的"集体民意",还要报告整个样本【Sample】的平均量表分——通常与另一个在不同时间点从同一个总体【Population】中获得的样本作比较,或与同一时间点从不同的总体中抽取的样本作比较。

比收集数据使用的方法更关键的是,要回答的研究问题【Research Question】或要检

验的理论【Theory】。原始数据必须多于(实际上少于)可从世界中得到的所有信息的简单集合。原始数据同沙堆有何种区别？数据必须依据一个目的而被辨识和收集。正是研究问题驱动着在所有可能的数据中收集其中某些子集,并排除其他信息。

最后,所有数据应该受到各种各样的评估,包括根据数据被收集的目的,判断数据中包含的误差程度。这类评估包括确定数据的信度【Reliability】和效度【Validity】,还包括给出舍弃数据的合理理由,以及对缺失数据【Missing Data】问题的考量。

百科全书中同样描述了数据的词条包括档案研究【Archival Research】、自传【Autobiography】、分类数据分析【Categorical Data】、截删数据【Censored Data】、计算机辅助数据收集【Computer-Assisted Data Collection】、横截面数据【Cross-Sectional Data】、数据【Data】、实验【Experiment】、网络调查【Internet Survey】、口述史【Oral History】、定群数据分析【Panel Data Analysis】、个人档案【Personal Documents】、二手数据【Secondary Data】、时间序列【Time Series】。

——John P. McIver

(王玥译 赵锋校)

参考文献

Coombs, C. H. (1964). *A theory of data*. New York: Wiley.

Diesing, P. (1991). *How does social science work?* Pittsburgh, PA: University of Pittsburgh Press.

Hyde, R. (2003). *The art of assembly language*. San Francisco: No Starch Press.

Page, B. I., & Shapiro, R. Y. (1992). *The rational public*. Chicago: University of Chicago Press.

RC(M)模型(RC[M] Model)

参见关联模型【Association Model】。

反应性测量(Reactive Measures)

参见反应性【Reactivity】。

反应性(Reactivity)

反应性这一术语,简单地指行为的观察或测量会影响行为本身,从而作出的解释会受误差的影响。可以将反应性问题理解为研究中的“人为要素(artifacts)”(Rosnow & Rosenthal, 1997),这一更普遍且有众多表现形式的难题的一个方面。例如,罗伯特·罗森塔尔(Rosenthal, 1966)观察到,实验员对女性实验对象微笑多于男性实验对象,他还颇具想象力地将其解释成“骑士精神在心理学实验中没有消失”。罗森塔尔警告说,实验员的微笑影响实验对象的反应,这在心理学上是有趣的,但它的方法论后果则令人

困惑。他将实验员导致的偏差效应归之于未控制的心理社会和生理社会的因素，同时，他也指出，非人类对象的实验研究也容易受到反应性效应的影响。例如，他陈述了这样一个发现，即在动物实验室中有经验的观察者能够根据老鼠跑迷宫或被抓起时的行为，猜测出哪几个实验员曾经摆弄过老鼠。

当研究中使用自我报告的人格测量法时，反应性问题有时表现为社会赞许性偏见【Social Desirability Bias】。这意味着，受访者不愿意显露他们真实的情感和信念，而是以他们认为能给他们带来好印象的方式作答。识别这类偏见的一个标准方法是，将个性测量的响应和社会赞许的测量作相关分析。目前，已有一系列关于自我报告测量的研究和理论，既包括检测反应性的程序，也包括控制的方法(Stone et al., 2000)。

另一种区分是反应性测量和非反应性测量的分别，后者指那些不影响行为的测量(Webb, Campbell, Schwartz, Sechrest, & Grove, 1981)。社会学研究者使用的各类非反应性测量方法包括：物理痕迹(physical traces)法、隐秘工具(covert instruments)法和故意隐踪观察(contrived unobtrusive observations)法。例如，韦布等人引证，一些研究者使用绘有婴幼儿的艺术品的照片来研究社会历史的相似与差异。虽然隐秘【Deception】工具(如隐蔽的麦克风或隐藏的相机)在被使用时经常引起伦理的担忧，但是，韦布等人主张多数社会研究中进行的隐秘测量并不足以产生危害，也不会引起严重的伦理问题。应用故意的、隐踪的观察法的研究，如在某些田野试验【Field Experiment】中，不仅使用隐藏的测量工具，还会使用一些操控(manipulation)工具。

——Ralph L. Rosnow

(马妍译　赵锋校)

参考文献

Rosenthal, R. (1966). *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts.

Rosnow, R. L., & Rosenthal, R. (1997). *People studying people: Artifacts and ethics in behavioral research*. New York: W. H. Freeman.

Stone, A. A., Turkkan, J. S., Bachrach, C. A., Jobe, J. B., Kurtzman, H. S., & Cain, V. S. (Eds.). (2000). *The science of self-report: Implications for research and practice*. Mahwah, NJ: Lawrence Erlbaum.

Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, L., & Grove, J. B. (1981). *Nonreactive measures in the social sciences* (2nd ed.). Boston: Houghton Mifflin.

实在论(Realism)

实在论者坚信某些值得怀疑的、实体的、实际的或可能的存在。这些实体可能是亚原子粒子、某一社会结构或世界及宇宙本身。作为一种普遍的哲学立场，实在论常对立于怀疑论——该学说认为我们无法知道事物(或任何东西)是否存在，也对立于如下立场，即把对实体的参照仅视作使得观察有意义的实用的虚构。实在论倾向于和唯物主义连同起来对抗唯心论【Idealism】，同时，就类词(general terms)的“普遍特征”或所指事物，实在论还对立于唯名论(nominalism)。在后一意义上，它与唯心论相接近，都断言某些东西的真实存在，比如说，与其在具体的红色对象中所显示的式样

有别,红性的真实存在。在现代科学哲学中,实在论主要拒斥实证主义【Positivism】,以及由"逻辑实证主义(logical positivism)"或"逻辑经验主义(logical empiricism)"提出的主张,即认为关于对象——好像能够脱离我们对它们的观察而实存——的形而上学问题,是无意义的。最近,随着实证主义(参见后经验主义【Postempiricism】)的没落,实在论者,特别是那些认同批判实在论【Critical Realism】的,已经把他们的反对点更多转向社会建构主义【Constructionism】和后现代主义【Postmodernism】。当然,也有其他实在论者把实在论的科学概念同说明人类社会的社会建构主义结合起来。

实在论的论证分为两大类别:第一类,是从科学的汇归(convergence)而来的归纳论证。如果科学家们对于经过多年精炼的解释逐渐地取得一致,就有理由认为我们正在接近于一个实在的可靠表象。第二类,是关于原则的论证,即为了使得科学成为可能或有意义所需立基于其上的原则——特别是,世界以一种有序的和可预测的方式存在。在社会科学中,两组论证都是问题重重:不仅不存在对一致认可的范式【Paradigms】的汇归,甚至就社会科学的可能性(参见自然主义【Naturalism】)都缺乏完全一致的认可。更可能的是,社会科学家在参照(现实的种种)社会结构和社会机制之前,已经有了一个含蓄的"好像是……"。

实在论并不将自己托付给任何关于社会世界的特定理论,或任何特定的研究方法。然而,如同"反经验主义(antiempiricist)"或"后经验主义(postempiricist)"的立场,它贬低了作为成功的解释标准的重复观察和预测的重要性。尽管理论分析和实验能表明实体的存在,但是那些只有在其效应中才能被观测的实体,如磁场和社会结构,却可能与其他实体互相影响,以至于效应相抵。比如,在地球的表面,就同时受到重力吸引的作用和地球旋转的离心力的作用;其结果是仍留在原地。社会力无论个人的还是结构的,都可能以相似的方式互相影响,可是它们却更难获得普遍满意的证明。反实在论者(antirealists)认为一种明显的不受控制的实体的激增是有问题的,而这些实体被赋予了像人一样的因能(causal powers)(参见因果关系【Causality】),同时,他们还责难实在论者的本质主义【Essentialism】。

——William Outhwaite

(王玥译　赵锋校)

参考文献

Outhwaite, W. (1987). *New philosophies of social science: Realism, hermeneutics and critical theory.* Basingstoke, UK: Macmillan.

Sayer, A. (1992). *Method in social science: A realist approach* (2nd ed.). London: Routledge.

Sayer, A. (2000). *Realism and social science.* London: Sage.

双向关系(Reciprocal Relationship)

双向关系是指一组假设的关系,其中某事件 X 影响第二个事件 Y,反之亦然。几项统计技术可用于判定双向关系是否存在:包括 F 比值【F Ratio】、间接最小二乘(indirect least squares, ILS)、工具变量【Instrumental Variables】、三阶最小二乘(three-stage least Squares,3SLS),以及二阶最小二乘【Two-Stage Least Squares】(也称作非递归因果模

型)。双向关系常出现在因果模型【Causal Modeling】或者结构方程估计中。

考察一下,工具变量法的一个特例,二阶最小二乘(2SLS)(Kennedy, 1998, p.165)。若要实施最小二乘,就要接连估计两个方程。第一个方程"以所有内生变量【Endogenus Variable】为回归项,建立其与联立方程系统中所有外生变量【Exogenous Variable】的回归关系……进而计算这些内生变量的估计值"(Kennedy,1998,p.165)。第二个方程当"将这些估计值作为这些内生变量的工具变量,或者仅使用这些估计值,并将外生变量作为回归项,建立一个最小二乘【Ordinary Least Squares, OLS】方程"(Kennedy, 1998, p. 165)。此方法先在一个联立方程系统中,逐个地将内生变量回归于外生变量,再将估计值当作回归项,带入最小二乘方程中计算,以此来检验双向关系的可能性。

一个双向关系的示例便于更充分地理解这个概念。政治科学家丹·伍德用一个例子说明了在他的环境政治学(environmental politics)研究中,双向关系如何发挥作用。伍德主张,涉及清洁空气的美国环境法在联邦层面的执行依赖于国家和地方两个层面对这些法律的执行(Wood, 1992)。通过使用工具变量法,即他设计的第一个方程预测了州层面的法律执行,他检验了他的观点。他的第二个方程以第一个方程所得的预测值为新的自变量【Independent Variable】,进而预测联邦层面对环境法的执行。在这个示例中,存在一个双向关系。因为州层面的法律执行依赖于美国联邦环保部对州项目的财政支持,同时,州的行动又影响了它们所获支持的程度(Wood, 1992, p. 50)。这个双向关系很重要,伍德(Wood,1992)必须明确两件事,以确定这一双向关系的存在:一是联邦财政支持的水平对州执行水平的效应;二是州执行的水平对每个州能获得多少联邦财政支持的效应。伍德用工具变量法得到的结果推论,在州执行水平和联邦环保部给予的财政支持度间存在双向关系。

——Kenneth W. Moffet

(马妍译　赵锋校)

参考文献

Kennedy, P. (1998). *A guide to econometrics* (4th ed.). Cambridge: MIT Press.

Kmenta, J. (1997). *Elements of econometrics* (2nd ed.). Ann Arbor: University of Michigan Press.

Wood, B. D. (1992). Modeling federal implementation as a system: The clean air case. *American Journal of Political Science*, *36*, 40-67.

再编码(Recode)

再编码一词,描述了改变变量【Variable】的值的过程。在收到或集齐一份数据集后,研究者不可能将其直接进行数据分析。通常必须先改变变量的值,才能用于分析。再编码一个变量的原因有很多,罗列如下:

1.数据的原始编码【Coding】存在必须改正的错误。就某一变量,计算其频次分布【Frequency Distribution】或描述性统计量,这可以帮助研究者查出必须通过再编码予以纠正的潜在错误。

2.通常收集到的数据中总是包含了尽

可能多的信息,但是必须再编码这些数据才能获得更具解释力的结果。如在调查中,回答人可能按要求报告他们的出生日期。将这些值再编码成年龄(以年为单位),研究者就能以更直观的方式解释该变量。

3.如果一个变量中某些值的编码不便于分析,也必须再编码。如在调查中,回答"不知道"和"拒绝回答"常被编码为8和9。研究者需要再编码这些值,使得所用的计算机软件程序能将其识别为缺失值。保持这些值不变将产生错误的结果。

4.给变量再编码的另一个常见原因是变原编码成序数的编码。在调查中,回答人的党派认同可能被编码为1=民主党,3=共和党,5=独立;可以重新将其编码为1=民主党,2=独立,3=民主党,由此编码就变成序数的编码。

5.还可以用再编码把一个变量的值归并成较少的类别。如回答人的收入可能是以10 000为区间分类记录的,其中,1=$0到$10 000,2=$10 001到$20 000,3=$20 001到$30 000,依次下去,共计10个类别。现在可以将收入再编码为低收入、中等收入、高收入,即将原来的编码1,2,3再编码为新的值1(低收入),将原编码4,5,6再编码为新的值2(中等收入),以此类推。

6.再编码某个变量的另一个原因是为了显出两个变量间的真实关系而变换该变量的值(如一个对数变换)。

7.此外,如果在原来的编码过程中,给缺失值赋了一个数值,那么研究者需要在分析数据前再编码这些缺失值。建议对再编码的变量计算一次它的频次或描述性统计量,以确保再编码符合要求。

——Charles Tien

(王玥译　赵锋校)

*也可参见编码【Coding】。

参考文献

Nagler, J. (1995). Coding style and good computing practices. *Political Methodologist*, *6*, 2-8.

Norusis, M. J. (1990). *SPSS base system user's guide*. Chicago: SPSS.

记录-核实研究(Record-Check Studies)

记录-核实研究通过两个来源获取同一个事实的数据,通常一个来源是家庭调查,另一个来源是某种行政记录。很多时候,其目标在于评估家庭调查的测量误差;这类研究也常被称作确认研究(validation studies),一般还假定记录资料是没有错误的。有时,在计算调查的估计值时,还用记录的数据代替家庭调查报告中的数据。如马克林和泰勒(Machlin & Taylor, 2000)所描述的,医疗支出追踪调查(the Medical Expenditure Panel Survey)中的医疗供应者调查(the Medical Provider Survey)部分就是一个数据置换记录核查的例子。

马奎斯(Marquis, 1984)和格罗夫斯(Groves, 1989)各自提供了确认型记录-核查研究的具体定义。马奎斯以一个二乘二的矩阵,根据一个二分变量的值(例如,某种慢性病状况的存在或不存在),为记录核查

研究划定了一个基本类别。就同一疾病状况，两个数据源可均予以肯定（矩阵的 A 格）；两者均否定（矩阵的 D 格）；或者记录为肯定，调查为否定（格 B）；抑或调查为肯定，记录为否定（格 C）。一种完全设计（full-design）的记录核查要容许所有四个格子的定量测量。在某些“对角记录核查（reverse record check）”研究中，仅当调查回答人报告肯定的答案时，才对记录予以核查（即一个“AC”设计）。在某些“前向记录核查（forward record check）”研究中，仅对那些在记录中肯定地确认的个体才进行调查（即一种“AB”设计）。

一个完全设计的记录-核查研究的例子是由爱德华兹等人（Edwards et al.，1994）给出的。他们在研究中使用了美国国家家庭访谈调查（National Household Interview Survey，NHIS）问卷，调查了从一个职员型健康维护组织（Health Maintenance Organization，HMO）的成员中选取的回答人。他们用 kappa 统计量报告了两个来源数据的一致度，同时发现，在某些慢性病状况（如糖尿病和高血压）上，数据有极佳的一致度，而在另一些慢性病状况（例如慢性支气管炎、心脏杂音）上，数据的一致度很差。他们还比较了根据每一来源计算的总体估计值，发现家庭调查的估计值同得自记录的估计值相比，即存在高报（例如心脏杂音和耳鸣），也存在低报（例如皮炎和心脏病）。

格罗夫斯发现确认型记录核查的两个限制：（a）假定记录是无误的，而事实上，记录系统经常会出现同调查同类的错误；而且（b）在比较两个数据源之前，必须将家庭报告的数据同记录数据匹配起来，而这是另一个容易导致疑问和错误的过程。例如，在爱德华兹等人的研究中，一种“真的”疾病状况可能没有记录，或者因为病人从没有提及它，或者因为在病人加入 HMO 之前已经得到了诊治，或者由于医务人员的失误。如马克林和泰勒就 MEPS 所做的描述，由于匹配特征时会发生错误，如病人的姓名、事件的数据、事件的类型，还有事件的缘由，因此匹配医疗事件是受概率支配的过程。

——W. Sherman Edwards

（王玥译　赵锋校）

参考文献

Edwards, W. S., Winn, D. M., Kurlantzick, V. et al. (1994). *Evaluation of National Health Interview Survey diagnostic reporting*. National Center for Health Statistics. Vital Health Stat 2(120).

Groves, R. M. (1989). *Survey errors and survey costs*. New York: Wiley.

Machlin, S. R., & Taylor, A. K. (2000). *Design, methods, and field results of the 1996 Medical Expenditure Panel Survey Medical Provider Component* (MEPS Methodology Report No. 9, AHRQ Pub. No. 00-0028). Rockville, MD: Agency for Healthcare Research and Quality.

Marquis, K. (1984). Record checks for sample surveys. In T. Jabine, E. Loftus, M. Straf et al. (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines*. Washington, DC: National Academy Press.

递归的(Recursive)

许多社会科学理论【Theory】设定的模型含有多个方程。一个递归方程组是联立方程【Simultaneous Equation】组的一类。递归方程组由因变量的因果链结合在一起。在递归模型中,(a)模型中影响的方向是单向的(没有双向因果或者反馈回路);(b)方程的误差项不相关。若一个方程组被认作是递归的,两个条件就必须同时得到满足。

考察下面的递归方程组:

$$y_1 = \gamma_{11}x_1 + u_1$$
$$y_2 = \beta_{21}y_1 + \gamma_{21}x_1 + \gamma_{22}x_2 + u_2$$
$$y_3 = \beta_{31}y_1 + \beta_{32}y_2 + \gamma_{32}x_2 + u_3$$

其中,$\mathrm{Cov}(u_1, u_2) = \mathrm{Cov}(u_1, u_3) = \mathrm{Cov}(u_2, u_3) = 0$。

在方程组中,内生变量【Endogenous Variable】标作 y,外生变量【Exogenous Variable】标作 x。β 表示一个内生变量对另一个内生变量的影响,同时,γ 表示一个外生变量对一个内生变量的影响。方程中的误差由 u 表示。

模型可以用路径图【Path Diagram】画出来,以图示的方法呈现出变量间的关系(图1)。

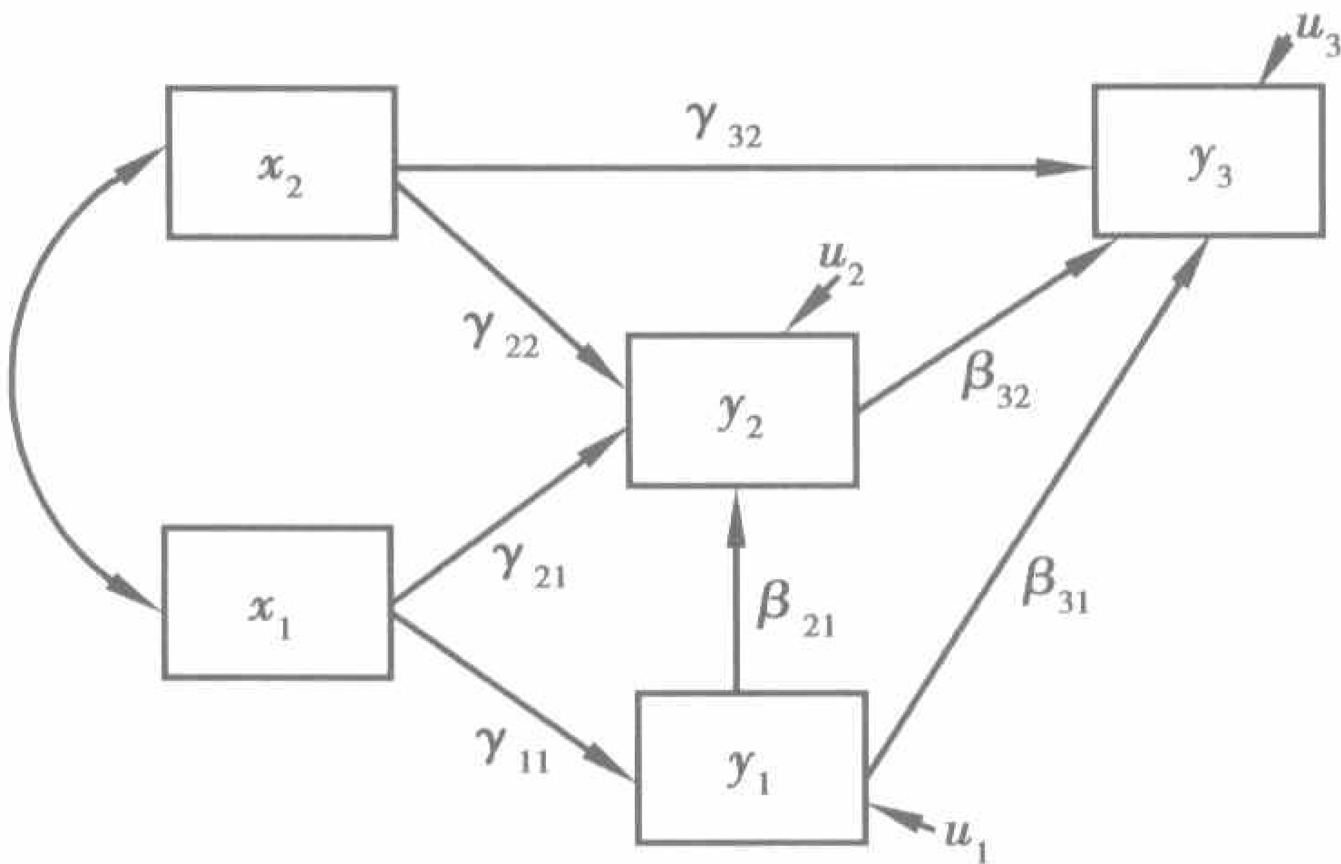


图 1

模型【Model】也可以用系数矩阵【Matrices】表示。注意方程中的下角标表示矩阵中系数的位置。

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} + \begin{bmatrix} \gamma_{11} & 0 \\ \gamma_{21} & \gamma_{22} \\ 0 & \gamma_{32} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

如所有递归方程组一样,可将上文的模型看成因果的序列(y_1影响 y_2,y_1和 y_2影响 y_3),并且一个方程的扰动项与另一个方程的扰动项不相关。在递归模型中,形式上,B 可以写成下三角矩阵(系数只出现在对角线以下的位置),方程中误差(u)的方差-协方差矩阵是一个对角阵。如果模型中有任何反向的因果关系,如 y_3对 y_1的影响(直接的或是通过其他变量链),或者,如果任意一个或所有的误差项相关,该模型就是非递归的【Nonrecursive】。

一个递归模型中的每一个方程都可以分别用最小二乘【Ordinary Least Squares, OLS】回归予以估计。

——Pamela Paxton

(马妍译 赵锋校)

*也可参见结构方程模型【Structural Equation Modeling】。

参考文献

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Greene, W. H. (1997). *Econometric analysis* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Heise, D. R. (1975). *Causal analysis*. New York: Wiley.

简化论(Reductionism)

在科学哲学中,简化论这一词一般指下述观念,即复杂的解释可以或应当化约成简单的。例如,生物现象可以由它们的化学成分来解释,而其化学成分转而可以由它们的原子或亚原子属性来解释。在社会科学中,简化论在不同的层次上发生作用,但遵循着相同的原则,也寻求复杂现象的简单解释。这是一个富有争议的策略,同时,有些人从事于此,却极少用这个词,而偏好(例如)方法论个体主义【Methodological Individualism】这样的表述。

简化论通常是比个体主义更加极端的策略,可以用于范围更广的现象。简化论曾是众多社会科学追寻的目标,特别是经济学、政治学、人类学和社会学。简化论的维多利亚式热情,即将社会现象的解释化约成个体的、遗传的生物特征在当代的复兴,可能作为这一理论的一个范例。它的现代表述可在进化心理学学科中找到。在多大程度上个体的生物特征决定了他们的社会特征和形式,这在各种进化理论的拥护者间有很大的差别,但是,最常采用的解释则依赖于基因遗传。

这一理论有两个要点——其一是遗传学,它越出了公认的观点,即认为我们的基因遗传预先影响了某些生物特征(疾病,如癌症),进而认为,我们的心理特征(如智力)或特定的行为模式(如越轨、空间能力、社交能力等),全部地或部分地归因于我们的基因。其二,进化心理学家主张人类基因组源自人类物种的自然选择。达尔文的自然选择原理排除了习得特征向后代的传递;只有繁衍的传递可以传递产生生活和行为特征的基因。自开化以来,人类在生理结构上少有变化,特别在脑的能力和特征上。因此,可以说,当大脑开始以不同于灵长类近亲的方式进化时,我们目前的基因禀赋就形成了。这一解释并不排除环境的交互作用,但是该理论认为,要想对人类行为作出解释,就必须先考虑开化前已定型的行为特征的基因遗传。

诸如此类的简化论解释已引发出一些非常有争议的社会研究。最著名的是赫恩斯坦和默里(Herrnstein & Murray,1994)的《钟形曲线》(*The Bell Curve*)。他们的核心论点是,美国大多数的社会和经济问题源于那些饱受贫困、失业和教育失败的人的低下智力。因此,他们推论,无论多少社会介入(例如 Head Start 项目)都不可能根本改变个体的生活机遇,也不可能改变这些个体生活的社会环境。他们的论证基于对全国青年追踪调查的 IQ 资料的分析,同时他们宣称,分析证明了 IQ 测验是生活成功的可靠的预测指标(Dickens, 2000, p. 67)。他们的研究既因为引出结论的相关关系的微弱而受到批评,也因为 IQ 的规范定义而受到批评。

更一般地说,对生物简化论的批评集中于支撑论点的数据缺陷,以及引用辅助假设解释经验不一致的倾向(Dupré, 2001, p. 68)。然而,对那些在社会科学中排除行为的所有生物学前提的人的挑战是,需要说明社会行为如何从生物学中独立出来。

——Malcolm Williams

(王玥译　赵锋校)

参考文献

Dickens, P. (2000). *Social Darwinism*. Buckingham, UK: Open University Press.

Dupré, J. (2001). *Human nature and the limits of*

science. Oxford, UK: Oxford University Press.
Herrnstein, R., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

反身性(Reflexivity)

反身性是一个有着不同用法的术语(Bourdieu & Waquant, 1992; Lynch, 2000)。对某些作者而言,例如,米德和加芬克尔,反身性是人类社会生活的中心特征,尽管他们两人给予这个词不同的意思。对米德而言,个体由于能采用他人的态度,以及由此能反思他或她的行为,而变成自我(参见符号互动论【Symbolic Interactionism】)。对加芬克尔而言,反身性指称如下事实,即社会行动在本质上是可说明的,在这个意义上,社会行动即是在行动之中且通过行动所展示的东西,也是行动的参与者所理解的东西(参见常人方法论【Ethnomethodology】)。在此,我们主要关注反身性的另一种用法,即社会研究实践的一项特征。它既是事实,也可能且应当成为一种品行。一种观点认为,研究者总是他们所研究的社会世界的一部分;他们不可能越出其中,以获得超然世外的视角(Olympian perspective),或步出其外,从而达到一种"自无而观(view from nowhere)"的境地。沿着上述论点,研究者应当持续地反思他们自己在研究过程中的角色,以及研究过程所发生的广阔背景。这种反思应引导研究的进程;同样,在撰写他们的研究发现时,研究者应给出研究过程细致的说明,从而允许读者根据背景来评判他们的研究发现(Hammersley & Atkinson, 1995)。

人们提出了几项支持研究者反身性的论点,不过,其中某些观点与其他观点相互矛盾。最根本和最率直的观点认为,研究是一项复杂且困难的活动,不能被简化为遵循一套预先设定的规则。研究者不可能决定怎样做才最好,而是需要持续地反思已经做了什么、有多成功,以及未来怎样更好地继续。这是一个相对普遍认可的观念。这一论点的改进版是,为了记录和追踪反身性,并使得任何对研究发现的扭曲影响变得最小,研究者必须反思自己是如何影响被研究的情境和人。

关于反身性的另一个论点要求把社会学的理论应用于社会研究自身。一个例子是贝克尔将标签理论应用于研究偏倚的问题(Becker, 1967)。还有一个相当不同的例子是古尔德纳的"反身社会学(reflexive sociology)",即社会学被视作一种具有改造作用的社会运动,而这一运动持续地构建、应用并发展着对世界(以及它自身)的理解,从而促进人类集体目标的实现(Gouldner, 1973)。

对反身性的另一种解释与怀疑论的认识论【Epistemological】观念相联系。这一观点始于以下坚持,即知识从不只是存在事物的表象,而总是等同于一项社会建构,并由此反映了研究者的视角和社会位置。这一论点的几种版本认可如下观念,即通过反身性,研究可以成为一种透明的、产生真知的活动(Alvesson & Sköldberg, 2000)。对于其他人,反身性的优点完全存在于它持续地颠覆了所有倾向于相信说明可以表现世界的观点。这里,反身性等同于一项伦理的或政治的承诺,持续地提醒读者,所有的说明,甚至包括那些建构主义研究者的,都是各种人为的构造而不是表象(Woolgar, 1988)。根据这种立场,说明不能依效度而评判,至少不能被视作与实在相符合,反而必须根据政治的、伦理的或审美的标准来评估。实际上,首要的标准可能只是研究者反思自身的

程度。

尽管对反身性的思考有不同的方式，但是根据以上所有的论点，反身性成为研究中的关键品行，而且，在某些情况下，也是唯一可有的。然而，不仅对这种观念存在着批评，而且在那些倡导反身性的不同解释的学者间，甚至自称以同一方法进行反思研究的研究者间，也存在着争论（Czyzewski, 1994）。一些批评同时来自经验主义和反经验主义两个方向，例如常人方法学（Lynch, 2000）。一方面，批评说，对于研究过程的反思本身不会增加研究发现的预期效度。另一方面，某些常人方法学者批评反身性的观念，因为它本身暗含着研究者多少可以超越这个被反身地构建的世界，而研究者和他们所研究的事物其实都必然是其中的一分子。

还有的批评拒绝此观念，即反身性能够提供一个基础，而在此基础上，可以以一种方式，例如，古尔德纳提出的方式，给出某种对社会世界的单一、全面的理解（Hammersley, 1999）。对于要求研究者的反思式说明所产生的后果，有人也提出了担忧。有学者提出，这样做可能让年轻的研究者处于被监控的状态，等同于处于一种被支配的形式中（Paechter, 1996; Troyna, 1994）。

由于术语反身性容许各种解释，因此，尽管其中特定的解释作为后实证主义研究方法论的重要组成，得到某些人强烈的提倡，它们的价值还是为他人所挑战。

——Martyn Hammersley

（王玥译　赵锋校）

参考文献

Alvesson, M., & Sköldberg, K. (2000). *Reflexive methodology*. London: Sage.

Becker, H. S. (1967). Whose side are we on? *Social Problems*, *14*, 239-247.

Bourdieu, P., & Waquant, L. J. D. (1992). *An invitation to reflexive sociology*. Chicago: University of Chicago Press.

Czyzewski, M. (1994). Reflexivity of actors versus reflexivity of accounts. *Theory, Culture and Society*, *11*, 161-168.

Gouldner, A. V. (1973). *For sociology*. Harmondsworth, UK: Penguin.

Hammersley, M. (1999). Sociology, what's it for? A critique of the grand conception. *Sociological Research Online*, *4*(3). Retrieved from http://www.socresonline.org.uk/socreson line/4/3/hammersley.html

Hammersley, M., & Atkinson, P. (1995). *Ethnography: Principles in practice*. London: Routledge.

Lynch, M. (2000). Against reflexivity as an academic virtue and source of privileged knowledge. *Theory, Culture and Society*, *17*(3), 26-54.

Paechter, C. (1996). Power, knowledge and the confessional in qualitative research. *Discourse: Studies in the Politics of Education*, *17*(1), 75-84.

Troyna, B. (1994). Reforms, research and being reflexive about being reflective. In D. Halpin & B. Troyna (Eds.), *Researching education policy*. London: Falmer.

Woolgar, S. (Ed.). (1988). *Knowledge and reflexivity*. London: Sage.

回归（Regression）

在非实验的社会科学中，回归是运用最广泛的数据分析技术。在一个回归模型中，一个因变量【Dependent Variable】（常标作*Y*）是一个或多个自变量【Independent

Variable】的函数（常标作 X_1，X_2 等）。自变量被假定解释或至少预测了因变量测量的现象。最简单的回归模型断言只存在一个自变量：$Y = a + b_1X_1 + e$。这个双变量回归模型表示变量 Y 是变量 X_1 的线性累加函数，再加上一个常数项（记作 a 的固定值）和一个误差项（记作 e）。主要兴趣在 X_1 对 Y 的影响，即记作 b_1 的回归系数。为了获得 a 和 b_1 的数值估计，运用最小二乘【Ordinary Least Squares，OLS】方法，用一条直线拟合 X_1 对 Y 的观测值。一个更复杂的回归模型断言存在两个自变量：$Y = a + b_1X_1 + b_2X_2 + e$。这个多元回归模型表示，$Y$ 是 X_1 和 X_2 的线性函数（加上一个常数项和一个误差项）。一个多元回归分析【Multiple Regression Analysis】方程总是有两个或多个自变量。在下文中，将用一个数据实例来审查双变量和多变量回归技术。然后讨论相关的假定、历史和最新的进展。

双变量回归

社会科学家通常需要检验两个变量之间的关系。为了举例说明，下面将展示一个经验的例子。想象一下教育社会学家们在一个中西部的小的大学中寻求相关信息，试图了解他们的学生在毕业后的收入。他们想了解很多事情，包括基本的人口学因素的影响，如学生家庭的社会经济背景。特别的，如果可能，他们试图建立如下联系，即父母的教育成就和子女后来的工作收入间的关系。为了收集关于这些和其他研究问题的必要数据，他们从 15 年前毕业的 350 名学生的校友名录中抽取了一个随机样本【Random Sample】。出于研究经费的考虑，他们在每 10 个人中抽取一个人，得到了一个样本量为 35 人的样本。表 1 包含了他们调查的部分数据。第一列是学生的编码【Code】；第二列是父母教育变量，以父母中教育程度较高的一方所完成的正规教育年限测量；第三列是学生的收入变量，以学生所报告的毕业 15 年后年收入（单位是千美元）测量；第四列是学生的性别，值 1 = 男性，0 = 女性。在通用统计软件【Statistical Package】中，这些数据列示见表 1。

表 1　三个变量的数据集

个案	变量 1	变量 2	变量 3
	教育年限	收入	性别
1	6	18	0
2	6	30	0
3	8	21	0
4	8	24	0
5	8	33	1
6	9	27	0
7	9	30	0
8	10	36	1
9	10	45	1
10	10	21	0
11	11	35	0
12	11	45	1
13	11	31	0
14	11	30	1
15	11	36	0
16	12	42	0
17	12	51	1
18	12	54	1
19	12	38	1
20	12	46	0
21	12	42	1
22	13	39	0
23	14	52	1
24	14	39	0
25	15	57	1
26	15	49	1

续表

个案	变量 1 教育年限	变量 2 收入	变量 3 性别
27	16	60	1
28	16	47	0
29	16	59	1
30	16	54	0
31	17	61	1
32	17	53	1
33	18	47	0
34	19	67	1
35	20	75	0

目前的研究问题是父母的教育程度（变量 X_1）是否有助于解释学生后来的收入（变量 Y）。核心假设【Hypothesis】是它们是正相关的。成长在父母教育程度较高的环境中，孩子有更多的学习和进步的机会及动力，在工作中也会获得更高的收入。数据分析是否支持这一假设？在双变量回归中，第一步是审查散点图【Scatterplot】，如图 1 所示。

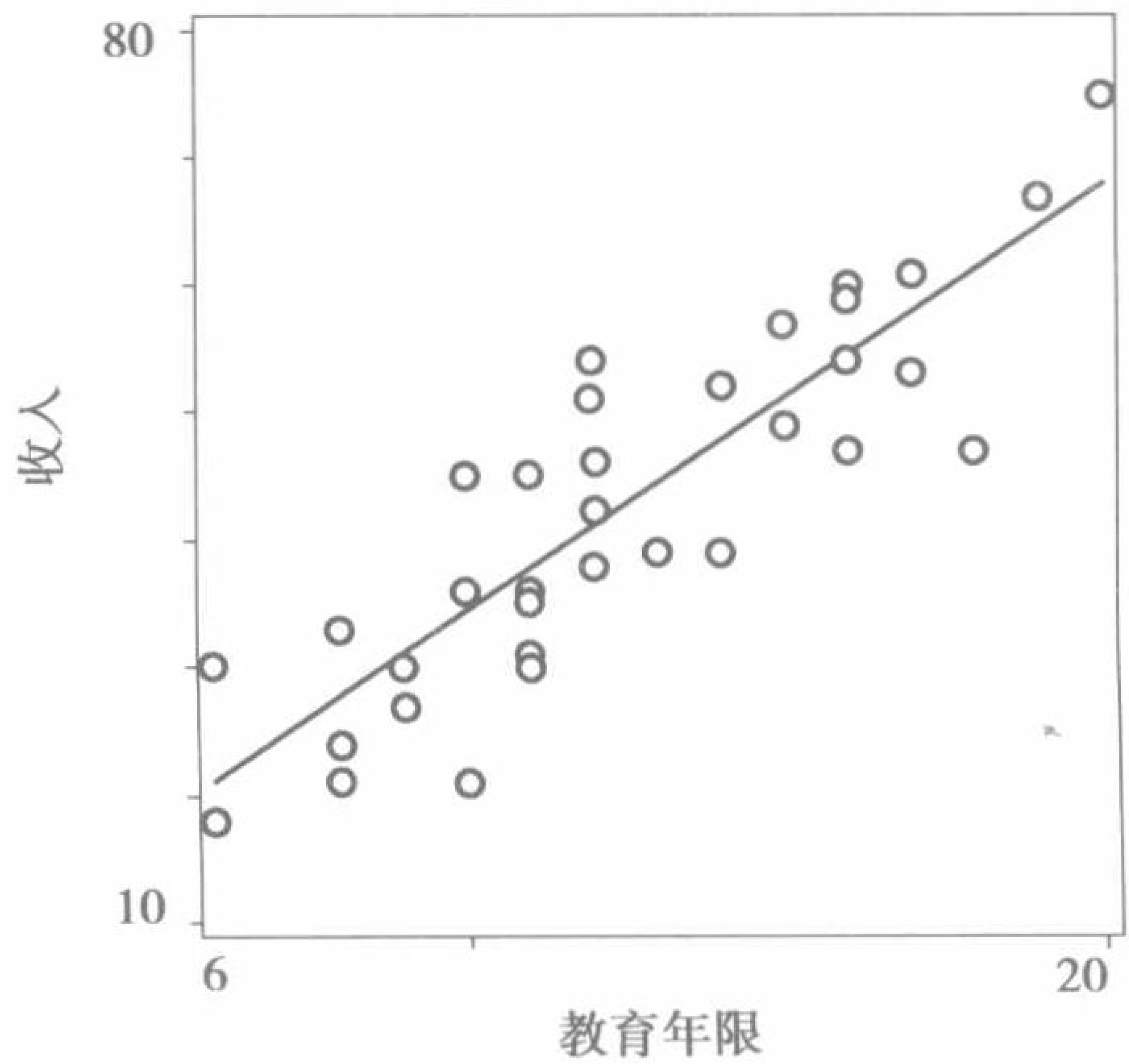


图 1　有回归线的散点图

图 1 中的每一个圆点或点代表一个特定的学生在变量 X_1 和 Y 上的取值。例如，第 35 号学生的父母教育程度取值为 20 年，自己的收入为 75 000 美元。右上角的点就代表她。实际上，代表每个学生的点位于从他或她的 Y 值（在纵坐标上）发出的假想垂直线和从他或她的 X 值（在横坐标上）发出的假想垂直线的交点上。对这些点的分散情况的审查显示 X_1 与 Y 正相关；也就是说，低的 X_1 取值趋向于有低的 Y 取值，出现在图的左下角；高的 X_1 取值趋向于有高的 Y 取值，出现在图的右上角。总之，点的陈列，即诸点呈现的几何形状，表明存在着一条直线。换言之，一条直线似乎比其他可能的曲线能触到更多的点。一条直线的公式是 $Y= a+ bX$。字母 a 是常数【Constant】，即直线与 y 轴的截距；而字母 b 表示直线的斜率。在图 1 中，给出了对散点图的实际拟合线，其回归方程为，预测的 $Y = 0.88 + 3.35X_1$。

图 1 中的回归线不是随意绘制的。相反，它是依据最小二乘【Least Squares】计算的最佳拟合线。可以观察到一部分点落在回归线上，其他的则没有。人们可以看到正的误差（落在回归线上）和负的误差（落在回归线下）。如果将所有这些误差（每个点到回归线的垂直距离）进行简单加总，由于正负（+和-）相消，它们的和为 0。然而，如果对这些误差取平方，就不会出现抵消，进而产生了总误差的一个量度，即这些误差的平方和（SSE）。这些误差的平方和是“最小的”，可能取得的最小值，亦即“最小二乘”表达的意思。根据微积分，可以计算使得 SEE 有最小值的截距（a）和斜率（b）的唯一值，而它们则构成了回归方程的最小二乘解。在该例中，其他截距【Intercept】和斜率的取值组合都不能像组合（0.88，3.35）一样产生一条对数据拟合得如此好的线。

回归线可以充当一个预测方程【Prediction Equation】。如果 Y 的预测值标

作 $\hat{Y}$,且 $X=12$,那么,

$$\begin{aligned}\hat{Y} &= 0.88 + 3.35X \\ &= 0.88 + 3.35 \times 12 \\ &= 0.88 + 40.2 \\ \hat{Y} &= 41.08\end{aligned}$$

模型预测出,如果学生父母的最高教育成就是 12 年,那么学生每年可以挣到 41 000美元。如果 $X=13$,那么

$$\begin{aligned}\hat{Y} &= 0.88 + 3.35 \times 13 \\ &= 0.88 + 43.55 \\ \hat{Y} &= 44.43\end{aligned}$$

人们可以发现,如果教育成就变量增加一年,收入的预测值相应增加 3.35 倍,即斜率值。这说明了斜率的一般性解释:b 表示一单位 X_1的变化伴随的 Y 的期望变化。截距也有一般性解释:a 表示当 $X_1=0$ 时 Y 的期望值【Expected Value】。当不存在 X_1为 0 的取值时,即如上述数据,那么截距只具有一个数学角色(它是为了完成预测所必须添加的一个恒量),而没有实际意义。

回归线可以用来预测,但是,正如那些远离回归线的数据点所表明,它们并不完美。线性模型对数据的实际拟合程度用 R^2 来衡量,而 R 方【R-Squared】估计了因变量上的变差可由自变量解释掉的比率。它是一个概括统计量,取值从 1 到 0。在该例中,$R^2=0.76$,表明父母的教育背景解释了学生收入上 76%的变差。

多元回归

多元回归分析允许方程中有多个自变量,从而扩展了简单线性回归。这是一项重要的扩展,有两个原因:首先,对一个现象的更复杂的解释成为可能,因为可以同时考察多种原因、预测因素或决定因素。其次,可以更好地理解特定的变量 X 对 Y 的具体效应,因为其他变量或者是混杂【Confounding】变量的影响可通过统计控制被排除。以前文的例子为例,除了父母教育程度之外的其他因素无疑有助于决定学生的收入。多元回归模型是为了将这些其他变量纳入考虑。这些其他变量之一是性别。于是,修正的理论是,学生的收入(Y)受到父母教育程度(X_1)和性别(X_2)的影响。以下是这一多元回归模型的最小二乘估计:

$$\hat{Y} = 0.44 + 3.12X_1 + 6.9X_2$$

根据回归结果,解释也需略加调整。同前面双变量的结果相比,父母教育程度每增加一年的影响似乎稍微减小了($3.12<3.35$)。斜率的减小可以归之于统计上"保持 X_2不变"。此外,性别自身似乎对收入有独立的影响。具体影响:如果控制父母教育程度的差异,男性能够期望比女性多挣 7 000美元。总体来说,$R^2=0.82$,表明这两个自变量共同解释了收入变量上 82%的变差。从双变量到多变量模型,R^2上升了六个百分点(调整的 R 方【Adjusted R-Squared】会稍微小一些)。

假定

对非实验的社会科学数据的回归分析旨在揭示真实世界中变量之间的结构性联系——因果的、解释性的和预测性的联系。更学术地说,目标是估计总体【Population】参数,即固定的截距值,以及始终同所研究的变量连在一起的斜率值。这些估计要精确到能够满足关键假定的程度。如果假定被忽略,回归结果就仅有纸上的价值,而无现实意义。若表 1 数据是真实的,那么给出的回归估计值是否揭示了学生收入、父母教育程度和性别之间的真实联系?是的,在某种程度上,它们得到了下面列出的假定支持。

绝大部分回归分析是根据一个样本【Sample】计算的,而不是根据作为整体的目标总体。基于样本方程的估计,如 $Y=a+b_1X_1+b_2X_2+e$,可以对总体方程中的参数进行推断,如 $Y=\alpha+\beta_1X_1+\beta_2X_2+\varepsilon$。样本必须依赖一个有足够规模的科学的概率抽样【Probability Sampling】。此外,由于样本只是样本,必须进行显著性检验【Significance Testing】,以排除机会结果。例如,根据从350名学生总体中抽选出35个样本,我们观察到多元回归方程中的斜率系数是3.12。对回归系数的显著性检验(0.05,双尾)可以让我们拒绝总体中父母教育程度与收入之间没有关系的虚无假设【Null Hypothesis】,亦即$\beta_1=0$的假设。

回归假定中除了有关于恰当抽样的假定以及与显著性检验有关的假定,还有许多经典的线性多元回归假定。大致而言,它们包括:没有设定【Specification】误差、没有测量误差、没有完全共线性【Collinearity】,且没有误差项问题。关于模型设定,根本的想法是模型具有正确的自变量,并且它们和因变量的关系是线性的。关于测量,要求变量应当是定量的,且有准确的赋值。关于共线性,则要求没有任何自变量是所有其他自变量的一个完全的线性函数。关于误差项,要求它作为方差同所有的自变量都不相关(以保证同方差性【Homoskedasticity】);作为一个变量(因为数据是非实验的数据),它应当同自身不相关(以避免自相关【Autocorrelation】)。当这些假定被满足时,OLS估计量是最佳线性无偏估计量【Best Linear Unbiased Estimator, BLUE】。某些分析者倾向于认为经典的线性回归假定形成了一个连续统,比如,从1~10的连续统,其中1表示完全没有满足,10表示完美地满足。从这个角度看,越接近10,推断就越接近现实。

旧的方向和新的进展

弗朗西斯·高尔顿(Francis Galton, 1822—1911)也许能被视为回归分析的创立者,因为他是第一个认识到可以用一条直线赋予一幅散点图上混乱的点以意义的人。他和他的弟子卡尔·皮尔逊(Pearson, 1857—1936)对父亲和儿子的特征的散点图作了大量考察。例如,皮尔逊考察了1 078对父亲和儿子的身高组合,发现很高的父亲倾向于有一个较矮的儿子。这就是"向均值回归",或者高尔顿称为"向平常回归"。描述这一回归效应的直线被命名为"回归线"。现代回归和相关的方法发端于《生物计量》(*Biometrika*)杂志上发表的关于遗传学的论文,特别是皮尔逊在1903年发表的成果。

OLS曾是回归估计的核心方法。最小二乘是由法国数学家阿德里安-马里·勒让德(Adrien Marie Lefendre, 1752—1833)和德国数学家卡尔·高斯(Carl Friedrich Gauss, 1777—1855)各自独立发现的。当经典的多元回归假定被满足时,OLS估计量无法加以改进。此后,特别是从1960年代开始,大量的工作涉及回归诊断,即试图检验假定是否得到满足,同时给出补救的方法。然而,某些模型本质上是非线性的,而且它们的参数不能用OLS予以有效估计。其中一种情形是:因变量是二分类变量,如"是"和"否"的投票选择。对此,一般倾向于最大似然估计【Maximum Likelihood Estimation】,如在Logistic回归【Logistic Regression】中。Logistic回归方法,包括多分类【Polytomous】的Logistic回归,曾是一个备受关注的研究领域。尽管这些方法在某些方面与OLS迥异,但是无论建模的逻辑,还是由于违背假定而产生的许多推断问题,都是相同的。

——Michael S. Lewis-Beck

(马妍译　赵锋校)

参考文献

Kmenta, J. (1997). *Elements of econometrics* (2nd

ed.). Ann Arbor: University of Michigan Press.
Lewis-Beck, M. S. (1980). *Applied regression: An introduction*. Beverly Hills, CA: Sage.
Neter, J., Kutner, M., Nachtsheim, C., & Wasserman, W. W. (1996). *Applied linear regression models*. New York: Irwin.

回归系数(Regression Coefficient)

回归【Regression】是一组普遍应用的方法,其目标在于用一组 k 个自变量【Independent Variables】(X_k)解释一个因变量【Dependent Variable】(Y)的变异。在最小二乘【Ordinary Least Squares,OLS】的前提下,回归系数($\hat{\beta}_k$)不仅与自变量($X_1, X_2, \cdots, X_k$)一一关联,还给出了在保持其他自变量不变的情况下,相应的自变量的一个单位的变化对因变量的期望的直接效应。回归系数用因变量的单位来表示。此外,截距项($\hat{\beta}_0$)通常也被估计,而它是所有自变量取值为 0 时,因变量的期望值。

在有两个自变量(X_1和 X_2)和 n 个观测值的 OLS 估计中,计算截距和回归系数的公式如下:

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X}_1 - \hat{\beta}_2\overline{X}_2$$

$$\hat{\beta}_i = \frac{\left(\sum_{i=1}^{n}(y_i-\bar{y})(x_{1i}-\bar{x}_1)\right)\left(\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2\right)-\left(\sum_{i=1}^{n}(y_i-\bar{y})(x_{2i}-\bar{x}_2)\right)\left(\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)\right)}{\left(\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\right)\left(\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2\right)-\left(\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)\right)^2}$$

考察下面的假想案例:

表 1 回归系数的计算

I	y	x_1	x_2	
1	22	39	6	$\bar{y}=24.6$
2	21	40	22	$\bar{x}_1=35.2$
3	24	39	20	$\bar{x}_2=14.6$
4	24	34	8	$\sum_{i=1}^{n}(y_i-\bar{y})(x_{1i}-\bar{x}_1) = -82.1$
5	26	33	21	$\sum_{i=1}^{n}(y_i-\bar{y})(x_{2i}-\bar{x}_2) = 21.2$
6	32	27	20	$\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2) = -31.1$
7	20	31	16	$\sum_{i=1}^{n}[(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)]^2 = 967.9$
8	27	32	8	$\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2 = 197.6$
9	25	41	10	$\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2 = 338.2$

根据这些数据和上面的公式,我们可以得到如下的模型估计:

$$\hat{y} = 38.7 - 0.412x_1 + 0.025x_2$$

截距项表示,如果 x_1 和 x_2 都为 0,我们可以预测 Y 等于 38.7。在实际研究中,这个值或许有实质的意义,也可能没有。x_1 的系数表示,在保持 x_2 不变的前提下,变量 x_1 上一个单位的变化,可以预计因变量将减少 0.412个单位。同样,变量 x_2 的系数表示,当保持 x_1 不变时,它的一个单位的变化,可以预计因变量将增加 0.025 个单位。如果 OLS 模型的所有假定【Assumptions】都没有被违反,系数的估值是无偏的【Unbiased】,也是对真实总体参数的最大有效估计(maximally efficient estimates)。此外,所有的回归系数都有与之相应的标准误【Standard Error】,因此,可据此检验它们统计上不为零的统计假设。

上例显示了回归系数背后的直观含义,还展示了如何计算和解释一个简单模型的回归系数。在更复杂的模型中,回归系数的计算和解释都有所不同,例如,带有交互项【Interactions】的模型,或自变量之间有非线性【Nonlinear】关系的模型,或带有受限因变量(limited dependent variable)的模型。当模型的因变量为受限因变量时,回归系数通常由最大似然估计【Maximum Likelihood Estimation】加以计算,也不能将它们解释成对因变量的线性效应(Greene, 2003)。然而,基本的直觉是一致的——一个回归系数告诉我们,一个变量,在保持其他解释变量不变的前提下,如何作用于另一个变量。

——Kevin J. Sweeney

(马妍译　赵锋校)

参考文献

Achen, C. H. (1982). *Interpreting and using regression*. Beverly Hills, CA: Sage.

Greene, W. H. (2003). *Econometric analysis* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Gujarati, D. N. (1995). *Basic econometrics* (3rd ed.). New York: McGraw-Hill.

回归诊断(Regression Diagnostics)

常见的违反最小二乘【Least Squares】回归假定的事项包括:非线性、异方差性【Heteroskedasticity】,以及非正态分布的误差项。异常观测值也可能对回归方程的估值有过度的影响,同时,多重共线性【Multicollinearity】可能产生不稳定的估计。如果这些问题未经矫正,OLS 的估计值就不再是最佳线性无偏估计量【Best Linear Unbiased Estimators】,所以实施模型诊断十分必要。

在简单回归中,非线性能较容易地通过检视因变量 Y 对自变量 X 的散点图【Scatterplot】来评估。然而,在多元回归中,简单的散点图就不再适用。因为,简单的散点图显示的只是变量 Y 和第 s 个自变量 X 之间的边缘关系,而模型给出的则是偏回归系数【Partial Regression Coefficients】。因此,在多变量的条件下,常用偏残差图[partial residual plots,也称成分残差图(component-plus-residual plots)]评估模型的线性。第 j 个自变量的偏残差 $E(_j)$ 只是因变量 Y 对自变量 X_j 的偏回归加最小二乘残差的线性组合。绘制 $E(_j)$ 对 X_j 的散点图,再以简单回归的散点图同样的方式判定模

型的线性。

非线性关系能够通过变换 Y 和(或)变换 X(非线性是简单且单调的),也可用多项式方程【Polynomial Equations】(关系是非单调的,但在方向上仅有较少的变化),或者用局部回归【Local Regression】(非线性关系非常复杂),予以解决。

异常观测值是一些游离值【Outliers】(如给定的 X 值对应一个异常的 Y 值),并且具有大的杠杆作用(以异常的 X 值为特征),能对回归模型的系数产生过度的影响。

游离值的统计检验通常基于学生化残差(studentized residuals)(有时也称为标准化残差)。"均值漂移(mean shift)"回归模型,只包括一个虚拟变量【Dummy Variable】,以 1 作为异常观测值的编码,0 作为其他观测值的编码,是一种用于识别某个游离值的学生化残差的常用方法。如果虚拟变量的系数具有统计显著性,那么,观测值显然偏离了其余数据的模式。然而,因为选取的是最远端的游离值,而不是随机选择的观测值,所以单独的 t 检验是不恰当的。p 值的 Bonferroni 校正【Bonferroni adjustment】可以弥补相应缺陷。最大游离值的 Bonferroni p 值的公式为 $p = 2np'$,其中,p'是 t 检验中自由度为 $n-k-2$ 的未调整 p 值。

杠杆率(leverage)由帽值(hat value)h_i 测定,该值反映出观测值 Y_i对预测 $\hat{Y}$ 的影响有多大。在简单回归中,帽值

$$h_i = \frac{1}{n} + \frac{(X_i - \overline{X})^2}{\sum_{j=1}^{n}(X_j - X)^2}$$[①]

取值范围为$\frac{1}{n}$~1,均值为$(k+1)/n$。若帽值接近平均帽值的两倍,就认为杠杆率是显著的。

诊断强影响点【Influential Cases】的最常用的方法是 Cook 氏距离(Cook's Distance)(或更简单地称作 Cook's D)。该方法度量了某个异常观测值对斜率系数的影响。Cook's D 的计算如下:

$$D_i = \frac{E_j^{*2}}{k+1} \times \frac{h_i}{1-h_i}$$

目前,尚无一致认可的 Cook's D 的临界值,但一条宽松的经验法则是,如果 Cook's D 的值大于 $4/(n-k-1)$,就应当做进一步的检查。

在识别异常个案时,用图形检视数据很有帮助。在简单回归的条件下,散点图很有效。在多元回归中,加变量图(added-variable plots,也称为偏回归图)可用于估计联合影响。此外,无论在哪种回归条件下,影像图(influence plots,也称气泡图【bubble plots】)——将帽值、学生化残差和 Cook's D 绘制在一张图上——都是非常有效的方法。

异常观测值可能反映了数据的错误编码,这时,可修正或删除相应的观测值。然而,一个游离值可能有实质上的意义,因此,可能决定将它从其余的数据中分离出来,单独地予以处理。许多游离值的出现可能说明某个重要的解释变量没有纳入模型。如果不存在去除游离值的强烈理由,那么可用稳健的【Robust】回归(可降低强影响数据的

① 原文的公式为 $h_i = \sum_{j=1}(X_j - \overline{X})^2$ 是错误的。——译者注

权重)代替 OLS 回归。

异方差性【Heteroskedasticity】通常可以使用残差图评估,即绘制残差 E_i(或者学生化残差,E_i^*)对 $\hat{Y}$(Y 的预测值)的图。如果所有的 Y 值都为正,则可以使用散布水平图(spread-level plot),即绘制 $\log|E_i^*|$(散布对数)对 $\log\hat{Y}$(水平对数)的图。将斜率为 B 的回归线同散布水平图相对照,会给出有稳定方差作用(variance-stabilizing)的,变量 Y 的幂变换值(power transformation)p,而 $p=1-B$。如果误差项方差与特定的 X 成比例,那么,加权最小二乘【Weighted Least Squares】可以有效地代替 OLS。

非正态性可以用分位数比较图(quantile comparison plots)进行诊断。该图绘制了学生化残差 E_i^* 的分布对单位正态分布的分位数,或对 t 分布的分位数。如果分布是不偏斜的,那么自变量同因变量的关系大致是线性的。峰态(modality)则可以用残差直方图或者残差的密度估计给出最佳评估。变量 Y 或变量 X 的变换均可用于修正重尾的误差分布,同时,模型重设(即纳入一个缺省的离散的预测变量)有时可以弥补多峰的问题。

多重共线性可通过两个自变量之间一个较高的配对相关系数(0.8 或更高)来识别。但是,在没有出现高的配对相关系数的情况下,多重共线性也可能存在,因为它反映的是模型中每个自变量 X 同所有其他自变量 X 间的多重相关。因此,更有效的方法是,考察每个自变量 X 对所有其他自变量 X 的回归所得到的 R^2($R^2=0.6$ 或者更高才有意义)。每个自变量的方差膨胀因子【Variance-Inflation Factors, VIF】,一个同 R^2 有关的指标,定义为 $\mathrm{VIF}=1/(1-R_j^2)$,给出了共线性对斜率【Slope】B_j 的精度影响。VIF 的平方根是标准误由于多重共线性的影响而膨胀的膨胀因子。

——Robert Andersen

(马妍译　赵锋校)

参考文献

Cleveland, W. S. (1993). *Visualizing data.* Summit, NJ: Hobart.

Cook, D. (1994). On the interpretation of regression plots. *Journal of the American Statistical Association*, *89*(425), 177-189.

Cook, D. R., & Weisberg, S. (1999). *Applied regression including computing and graphics.* New York: Wiley.

Fox, J. (1991). *Regression diagnostics* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-079). Newbury Park, CA: Sage.

Fox, J. (1997). *Applied regression analysis, linear models, and related methods.* Thousand Oaks, CA: Sage.

Fox, J. (2002). *An R- and S-PLUS companion to applied regression.* Thousand Oaks, CA: Sage.

定序数据的回归模型(Regression Models for Ordinal Data)

在社会科学中,定序层次变量(如调查问卷中的李克特量表【Likert Scale】,社会阶级的测量等)常会被用作因变量【Dependent Variable】。在回归【Regression】框架内,此类数据能用几种方式生成模型:(a)可以忽略类别的排序,使用二分的(binary)或多项

Logit【Multinomial Logit】(多项对数单位模型);(b)可以把类别当作定距尺度的变量,从而能够使用最小二乘【Ordinary Least Squares,OLS】;(c)使用专为定序因变量设计的模型。上述每种方法既有其优势,也有其局限。

第一种策略(忽视因变量的序次特征)。当因变量类别很少时,这种策略是很可行的。如果因变量的类别能够合理地压缩为两类而没有损失太多的信息,那么可以使用简单的二分 Logit 模型(二分对数单位模型)(Logistic 回归【Logistic Regression】)。对于这些模型而言,可以使用概率比【Odds Ratios】来确定类别之间的差异。然而,这种方法的不足在于,经过压缩分类,可能损失了关于因变量的重要信息。如果选择不压缩类别,虽然可以使用多项对数单位模型,但是这些模型是无效率的,并且随着因变量类别的增加,由于存在大量的被估参数,所以解释的难度也不断增加。

第二种策略。使用 OLS 的优势在于参数估计很容易理解(参见斜率【Slope】)。当因变量的分布近似于正态分布并且排序的类别确切地代表一个潜在连续分布时,使用 OLS 是合理的。一方面,如果变量具有很多类别,可通过直方图【Histogram】对其分布做粗略的估计。另一方面,如果变量只运用了很少的几个类别测量,变量就会表现出相当高的测量误差水平,从而不能合理地评估这一假定。如果预测是研究的一个主要目标,这类模型也可能存在问题,因为预测值可能会落在变量范围之外。然而,如果研究的主要目的是确定变量的关系,而不是预测,那么即使已知变量的测量值落在允许的范围之外,这就不一定成为问题,因为变量的测量总是存在误差。

由于我们通常不能假定定序变量各类别之间的距离是定距的,这就意味着它们无法满足 OLS 的准则(特别是误差项方差恒定的假定),因此,人们发展了专门用于定序数据的回归模型。这些模型包括定序对数单位模型(ordered logit model),即 Logistic 回归的一个一般化,以及定序概率单位分析【Probit Analysis】模型,即二分概率单位模型的一个一般化。在此,我们介绍最常用的定序对数单位模型:比例-概率比模型(proportional-odds model)。

对数单位(logits)可以直接包含一个因变量上各类别的序次,其结果是,由此得到的模型同多项对数单位模型(常用于无序次的类别因变量)相比,容许有更简单的解释,并且具有潜在的更大的统计效力。设想有一个潜在的连续变量 ξ,它是一组解释变量 X 加随机误差项的线性函数,即

$$\xi_i = \alpha + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i$$

假定 ξ 在这里代表对流产的态度,β 是一组回归斜率。还假定只有一个单一的 5 点李克特量表对流产的态度,即因变量 Y。在这个例子中,潜在的连续变量 ξ 被分成 5 个类别,$m = 5$,因此,有 $m-1 = 4$ 分界点,$\alpha_1 < \alpha_2 < \cdots < \alpha_{m-1}$:

$$Y_i = \begin{cases} 1 & \text{如果 } \xi_i \leqslant \alpha_1 \\ 2 & \text{如果 } \alpha_1 < \xi_i < \alpha_2 \\ \vdots & \\ m-1 & \text{如果 } \alpha_{m-2} < \xi_i \leqslant \alpha_{m-1} \\ m & \text{如果 } \alpha_{m-1} < \xi_i \end{cases}$$

进而,很容易发现 Y 的累积概率分布为:

$$\Pr(Y_i \leqslant j) = \Pr(\xi_i \leqslant \alpha_j)$$

$$= \Pr(\alpha + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i \leqslant \alpha_j)$$
$$= \Pr(\varepsilon_i \leqslant \alpha_j - \alpha - \beta_1 X_{i1} - \cdots - \beta_k X_{ik})$$

由此，可以发现定序对数单位模型【Logit Model】：

$$\text{logit}[\Pr(Y_i > j)] = \ln \frac{\Pr(Y_i > j)}{\Pr(Y_i \leqslant j)} = (\alpha - \alpha_j) + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}$$

这里的对数单位是累积的对数单位，其对比项是类别 j 以上的诸类别与类别 j 和 j 以下的诸类别。诸自变量 X 只有一个斜率系数，但是每一个类别有各自的截距项。第一个类别（或者最后一个类别，取决于软件）被设定为所有截距都与之相关的参照类。

定序对数单位模型的系数的解释与二分对数单位模型相似，不过，二分对数单位模型参照于一个单一的基线类别，在定序对数单位模型中，我们用每一个类别及其下的类别同所有序次高于它的类别相比较。换言之，对数单位告诉我们，Y 落于某个类别和其下类别的概率与处于该类别之上的概率的概率比的对数。

尽管比例-概率比模型由于简洁而富有魅力，但是务必谨慎地使用它。该模型假定，诸自变量的回归斜率相等（被称为比例-概率比假定或平行斜率假定），而这在实践中通常难以成立，特别是在大样本的情况下。检验是否满足这个假定的一个简易方法是，将该模型的拟合与多项对数单位模型的拟合相比较，因为后者可以被视为一种更一般的模型。两个模型背离真实模型的差异给出了一个广义的似然比【Likelihood Ratio】检验，该统计量呈卡方分布，其自由度【Degrees of Freedom】等于多项对数单位模型参数数目和比例-概率比模型参数数目之差。

——Robert Andersen

（马妍译　赵锋校）

参考文献

Agresti, A. (1984). *An introduction to categorical data analysis*. New York: Wiley.

Fox, J. (1997). *Applied regression analysis, linear models, and related methods*. Thousand Oaks, CA: Sage.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B, 42*, 109-142.

Menard, S. (2002). *Applied logistic regression analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-106). Thousand Oaks, CA: Sage.

回归于……（Regression on...）

短语"回归于"是指将因变量【Dependent Variable】向量（有时称为归依项）投射到由自变量【Independent Variable】（有时统称为回归量【Regressors】）结构的欧几里得子空间的行为。虽然这个短语涉及最小二乘的潜在几何形状，但也常用它来表示模型因变量上观测到的变化的自变量组。

例如，若一研究者的兴趣在于简单线性回归【Linear Regression】$Y = b_0 + b_1 X + e_1$，我们可说"Y 对 X 的回归"，而这句话的意思是，这一简单回归从系数集 $\{b_0, b_1\}$ 取得系数，亦

即，Y 向 X 构成的子空间的投射(projection)(在该例中是一条直线)由 b_0+b_1X 给定。由于 b_0 和 b_1 有不同的取值(比如 b'_0 和 b'_1)，Y 向 X 构成的子空间的投射(即 $b'_0+b'_1X$)也不同，所以存在着许多(无限的)可能的投射或回归。要确定哪个投射或回归是最佳的，研究者必须根据某个给定的标准评估"吻合(fit)"。一个常用的标准是最小二乘【Least Squares】，它表示当 $|Y-(b_0+b_1X)|^2$ 有最小值时，有选择 $\{b_0, b_1\}$ 的"最佳"投射。

在多元回归分析【Multiple Regression Analysis】的背景下，"回归于"指因变量向量向一组自变量构成的子空间的投射——而回归面【Regression Plane】确定了相应的面，即该面使得因变量观测值和它在回归面上的投射之间的差值的平方取最小值。

——Joshua D. Clinton
(马妍译　赵锋校)

参考文献

Davidson, R., & MacKinnon, J. G. (1993). *Estimation and inference in econometrics*. New York: Oxford University Press.

回归面(Regression Plane)

回归(超)平面代表因变量到由自变量集所围成的欧式子空间的投影。

运用最小二乘【Least Squares】准则的多元回归【Multiple Regression】旨在找到一个超面，该面能够使得因变量的观测值同因变量到该面的投影间的差值平方和取得极小值。这个能极小化此一差值的超面即称作回归(超)面。

一个例子可以帮助澄清这个概念。设想一位研究者对因变量 Z 与自变量 X 和 Y 之间的(线性)关系感兴趣。进而有如下方程：

$$Z = b_0 + b_1X + b_2Y + e$$

考虑一下被拟合数据的性质——每个观测值 i 确定了三维空间(X_i, Y_i, Z_i)中的一个点。直观地说，上例中，回归面即能够"最佳拟合"三维数据点集的二维平坦的表面——而"最佳拟合"意味着回归面能够极小化面的表面(代表给定 X 和 Y 时，Z 的预测值)和观测到的 Z 值的差值平方和。就该例而言，回归面确定了一个二维的平面，而该平面能极小化因变量 Z 和平面间的残差平方和。而一般而言，如果存在着 k 个自变量(不包括常数项)，回归超面是 k 维平坦的表面，而这 k 维表面代表了因变量 Y 到由 k 个自变量加因变量所围成的(k+1)维空间上的最佳线性投影。回归面是当 $k=2$ 时的特例。

回归系数【Regression Coefficients】确定了回归面在由各个自变量所定义的维度上的斜率。例如，在上文界定的三维空间中，b_0 表示当 $X = Y = 0$ 时，回归面在第三维(即 Z)上的"高度"；b_1 表示回归面在第一维度的斜率(即该面与 Z-X 面相交的角度)；而 b_2 表示回归面在第二维度的斜率(即该面与 Z-Y 面相交的角度)。回归面的方程是 $b_0+b_1X+b_2Y$。换言之，正如简单回归在于找到最佳拟合线，多元回归的一般目的在于找到最佳拟合面。

举例说明，图 1 中的面是以回归方程 $Z=0.3-0.5X+2Y$ 的特征绘制的。因为回归面描述了 Z 向 X 和 Y 的投影，所以对于(X, Y)的任意组合可能产生一个 Z 的预测

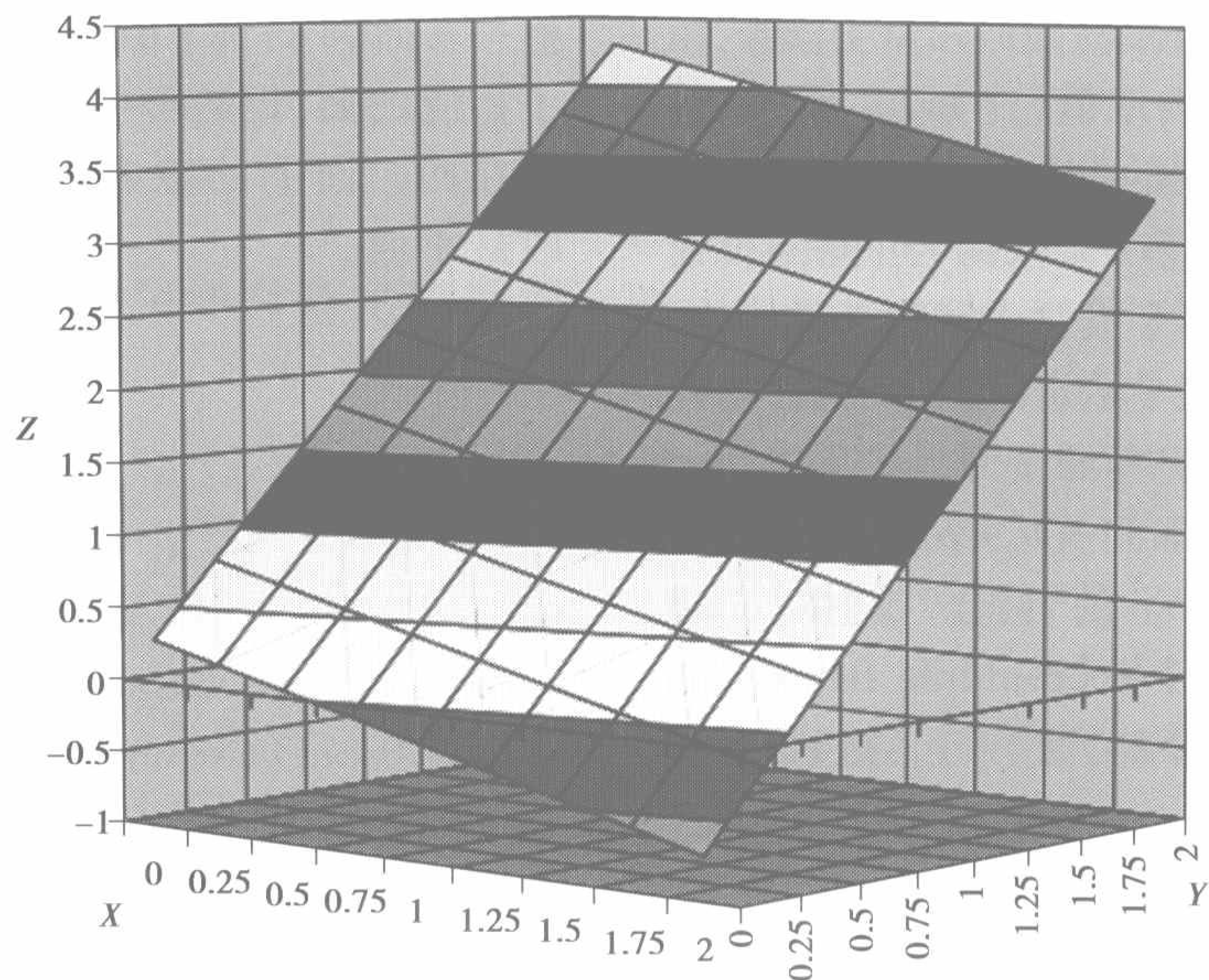


图 1 回归面

值。回归面即刻画了这些预测值。斜率系数的几何解释也能从图上得到清楚的呈现。X 的回归系数为-0.5，表示图中平面向左侧抬起；X 的取值越高导致 Z 的预测值越低。相比之下，平面在后部"抬升"因为 Y 的斜率系数为正。截距（在本例中是 0.3）由 $X=Y=0$时，平面"高度"给出。

——Joshua D. Clinton

（马妍译　赵锋校）

参考文献

Davidson, R., & MacKinnon, J. G. (1993). *Estimation and inference in econometrics.* New York: Oxford University Press.

回归平方和（Regression Sum of Squares）

在线性回归【Linear Regression】分析中，有必要报告回归线多大程度上拟合了数据。就线性回归模型而言，一个典型的拟合优度【Goodness of Fit】量度是响应变量（或因变量 Y）上的样本变差可以由自变量（X）上的变差解释掉的部分。回归平方和（RSS）是线性回归模型的总体拟合优度量度的关键部分。RSS 能告诉数据分析者变量 Y 上观测到的变差有多少可以归之于变量集 X 上的变差。换言之，RSS 表示变量集 X 对变量 Y 的变差的联合估计效应。

为了解释 RSS 的概念，就有必要界定因变量【Dependent Variable】Y 的样本【Sample】变差。对一个特定的观测值 Y_i，它在变量 Y 上观测到的变差即是它与均值的离差，或用符号表示为 $Y_i-\bar{Y}$。由于我们想要知道所有观测值的变差，一个合适的综括量度是所有观测值的离差的平方和。这就给出了变量 Y 的总变差或者总平方和（total sum of squares，TSS），即

$$\text{TSS(总变差 }Y) = \sum_i (Y_i - \bar{Y})^2$$

现在的问题是如何分解变量 Y 的总变差,亦即记作 TSS 的“总平方和”,由此我们才能度量变量 Y 的总变差的多大部分能归之于变量集 X 的变差。可以思考以下这个适用于所有观测值的等式:

$$Y_i - \bar{Y} = (Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y}) \quad (1)$$

其中,$\hat{Y}_i$ 是用一个等于 X_i 的自变量对变量 Y 的一个特定观测值的拟合值。就一个简单回归【Simple Regression】模型而言,即有 $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$。

已经定义了方程式等号左边的部分为一个特定的值 Y(即 Y_i)对 Y 的均值的离差。方程右边的第一部分给出了残差【Residual】e_i,因为,根据定义,$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i = \hat{Y}_i + e_i$,所以,$Y_i - \hat{Y}_i = e_i$。方程右边的第二项给出了 Y 的拟合值(或预测值)和 Y 的均值的差。由于 $\hat{Y}_i$ 是用自变量的值(X_i)预测得到的,右边的第二项即代表变量 Y 上的变差可以由变量 X 的变差解释掉的部分。一个简单回归(即只有一个自变量)中对变量 Y 的变差的分解展示在图 1 中。

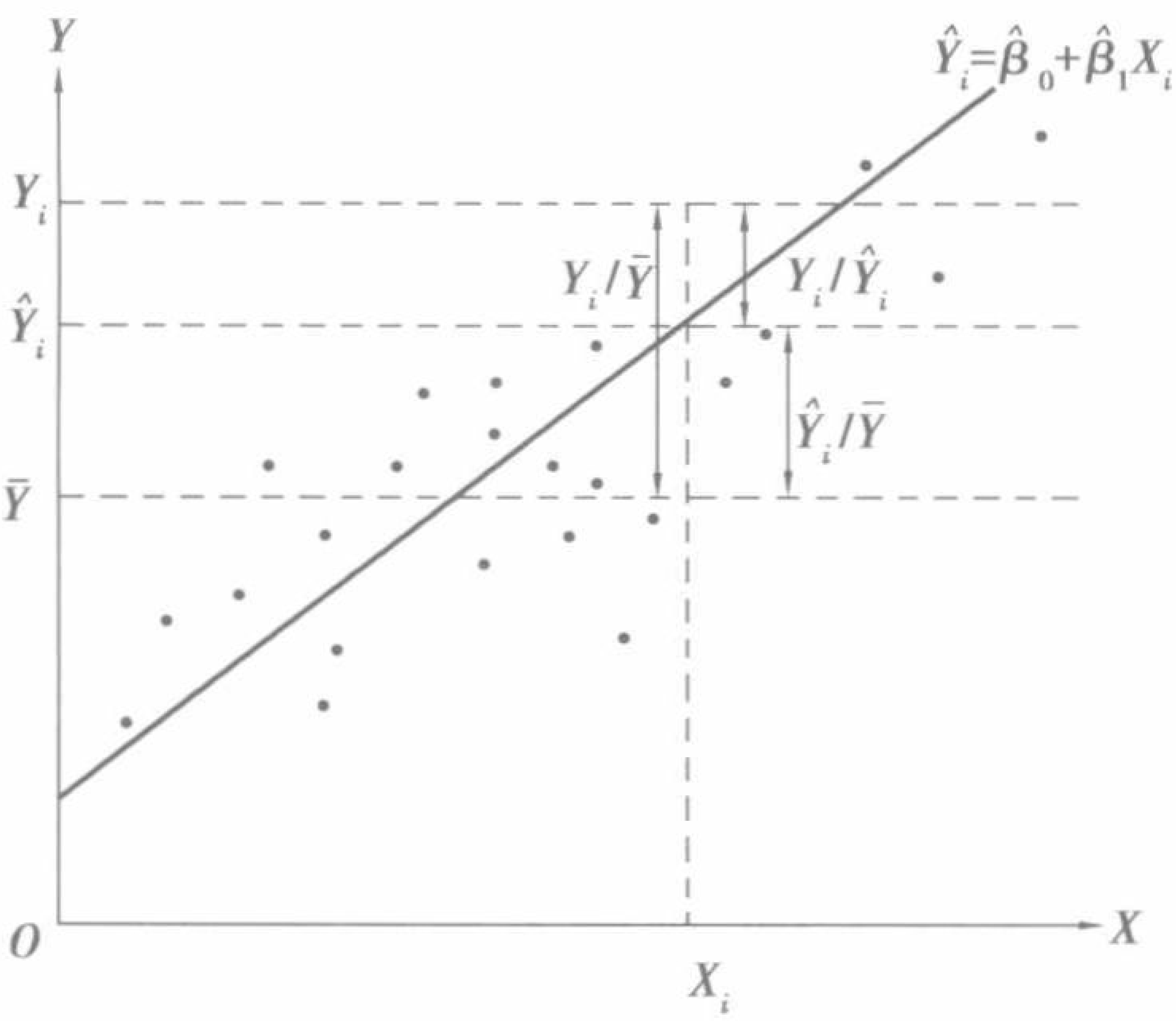


图 1 Y 的变差的分解

正如前文所述,方程(1)只适用于一个单一的观测值。对所有的数据,可以通过取等式两边的平方,再加总所有的样本观测值,从而得到一个综括量度。这就得出如下方程:

$$\sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 + 2\sum_{i=1}^{n} (Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) \quad (2)$$

基于经典正态线性回归【Regression】模型的两个假定【Assumption】:扰动项均值为 0[$E(e_i|X_i) = 0$];以及扰动项与自变量的协方差为 0 [$\text{cov}(e_i, X_i) = E(e_i X_i) = 0$],方程(2)的末项等于 0。进而方程(2)可以改写为:

$$\sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$$

总平方和 = 误差项平方和 + 回归平方和

(TSS) = (ESS) + (RSS) (3)

方程(3)表明,就所有观测值 $i = 1, 2, \cdots, n$ 而言,回归平方和(RSS)能被定义为因变量的预测值($\hat{Y}_i$)同因变量的均值($\bar{Y}$)的差的平方和。方程(3)也表明变量 Y 的变差一部分是由于随机扰动项的效应(即误差项平方和[ESS]),另一部分是由于变量 X 的变化(即回归平方和[RSS])。

变量 Y 上的总变差(或总平方和[TSS])可以归之于变量 X 的变差的比例是一个常用的拟合优度量度,也被称为(样本)决定系数【Coefficient of Determination】,并记作 R 方【R-Squared, R^2】。

注意,R^2 是一个非负的量,且不能大于 1(即 $0 \leqslant R^2 \leqslant 1$)。关于 R^2 其他属性的更多信息和关于怎样获得方程(3)的技术推导,详见福克斯等人的著作(Fox, 1997);Rawlings et al., 1999; Kmenta, 1986; Johnston &

Dinardo,1997)。

——Cheng-Lung Wang
(马妍译　赵锋校)

参考文献

Fox, J. (1997). *Applied regression analysis: Linear models and related methods*. Thousand Oaks, CA: Sage.

Johnston, J., & Dinardo, J. (1997). *Econometric methods* (4th ed.). New York: McGraw-Hill.

Kmenta, J. (1986). *Elements of econometrics* (2nd ed.). New York: Macmillan.

Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (1999). *Applied regression analysis: A research tool*. New York: Springer.

向均值回归(Regression Toward the Mean)

向均值回归(Regression Toward the Mean,RTM)是由弗朗西斯·高尔顿(Galton,1822—1911)首先关注到的,一种发生在任意两个不完全相关的变量之间的统计现象。这些变量可以是完全不同,但同时测量的两个变量(如身高和体重),也可以是在不同时点上测量的同一变量(时点 t 和时点 $t+1$)。无论是哪种情况,第一个变量上的极端观测值很可能在第二个变量上不那么极端,亦即更接近于均值。高尔顿的工作聚焦于豌豆种子和人类的遗传规模。他发现两类对象都“归向于普通”,即体格较大的父代有略矮小的子代,而体格较小的父代则有略高大的子代。如果忽略了 RTM 的存在,这就会成为实验的内在效度【Internal Validity】的一个威胁。

当两个变量不完全相关时($\rho<1$),RTM 就会出现。如果两个变量都是标准化【Standardized】的(两者都有均值=0,方差=1),且是正态分布【Normally Distributed】,我们可以使用它们的相关系数来估计期望的 RTM:

$$E(Z_Y|Z_X)=\rho Z_X$$

例如,如果母亲和女儿的体重都是标准化的,且她们体重的相关系数为 0.8,我们可以期望,一位体重高于“母亲”均值两个标准差【Standard Deviations】的母亲会有一位女儿,其体重将高于“女儿”均值 1.6 个标准差。重要的是,公式表明 RTM 的期望值会随着变量间相关系数的减小而增加。RTM 也是双向的:

$$E(Z_X|Z_Y)=\rho Z_Y$$

因此,可以期望,一个体重高于“女儿”均值两个标准差的女儿会有一个体重仅高于“母亲”均值 1.6 个标准差的母亲。

虽然上文给出的公式对估计一个个体 RTM 的期望值很有用,但 RTM 本质上是一种群体现象。任意一个特定个体在第二次观测时都可能远离均值,但是,只要两个变量间的相关关系是不完全的,群体将总是向均值回归。更进一步,一个群体越是极端,它就越是以绝对的条件向均值回归。群体均值接近于样本均值的群体并没有太多可回归的,而均值远离于样本均值的群体,就有很大的回归空间。这一点对许多社会科学实验的内在效度具有重要的隐含意义。通常,受试者的选取是因为他们有特别好或特别差的表现。特别是在后一种情形下,干预好似解决了问题,因为受试者总是显得有所改善。然而,如果该群体是极端的,改善将可以用 RTM 单独予以解释。

——Kevin J. Sweeney
（马妍译 赵锋校）

参考文献

Campbell, D. T., & Kelly, D. A. (1999). *A primer on regression artifacts*. New York: Guilford.

Galton, F. (1886). Regression toward mediocrity in hereditary stature. *Journal of the Anthropological Institute*, *15*, 246-263.

Smith, G. (1997, Winter). Do statistics test scores regress toward the mean? *Chance*, pp. 42-45.

回归量（Regressor）

回归量是一个描述回归方程中用以解释因变量【Dependent Variable】变差【Variation】的自变量【Independent Variable】的术语。回归量也被称为解释变量【Explanatory Variable】或者外生变量【Exogenous Variable】。在一个方程系统中，一个变量在某个方程中是回归量，可能在另一个方程中则是回归因变量，从而使得外生的（exogenous）这个术语的应用变得复杂。

一个回归量可能是一个定量变量【Quantitative Variable】（即以定距或定比尺度测量的变量），或者是一个定性变量【Qualitative Variable】（即具有名义或类别属性的观测变量）。在经典的线性回归模型中，为了方便起见，回归量被假定为非随机的或者"在重复样本中固定的"。这一假定可以放松，而不会产生太大的统计问题，特别是在样本规模较大的情况下。

——Brian M. Pollins
（马妍译 赵锋校）

参考文献

Greene, W. H. (2003). *Econometric analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Gujarati, D. N. (2003). *Basic econometrics* (4th ed.). Boston: McGraw-Hill.

Koutsoyiannis, A. (1978). *Theory of econometrics: An introductory exposition of econometric methods* (2nd ed.). London: Macmillan.

关系（Relationship）

关系是两个或多个概念【Concept】或变量【Variable】间的联系。例如，如果两个变量相关，那么它们就会以某种非随机、系统的方式共同变化。比如，随着 X 得分上升，Y 得分也上升。X 和 Y 间的这一联系可以用某种关联量度【Measure of Association】给予量的表示，至于使用哪一种关联量度则部分地取决于 X 和 Y 的测量层次。最常用的关联量度是皮尔逊相关系数【Pearson's Correlation Coefficient】，它可以衡量两个假定以定距尺度测量的变量 X 和 Y 之间的线性关系。一个关系也可能存在于三个或更多变量之间。例如，三个变量（X_1，X_2 和 X_3）都与变量 Y 相关。如果所有的变量都是定距的，那么可以计算多元相关【Multiple Correlation】系数 R 来估计这一关系。

——Michael S. Lewis-Beck
（马妍译 赵锋校）

相对分布法(Relative Distribution Method)

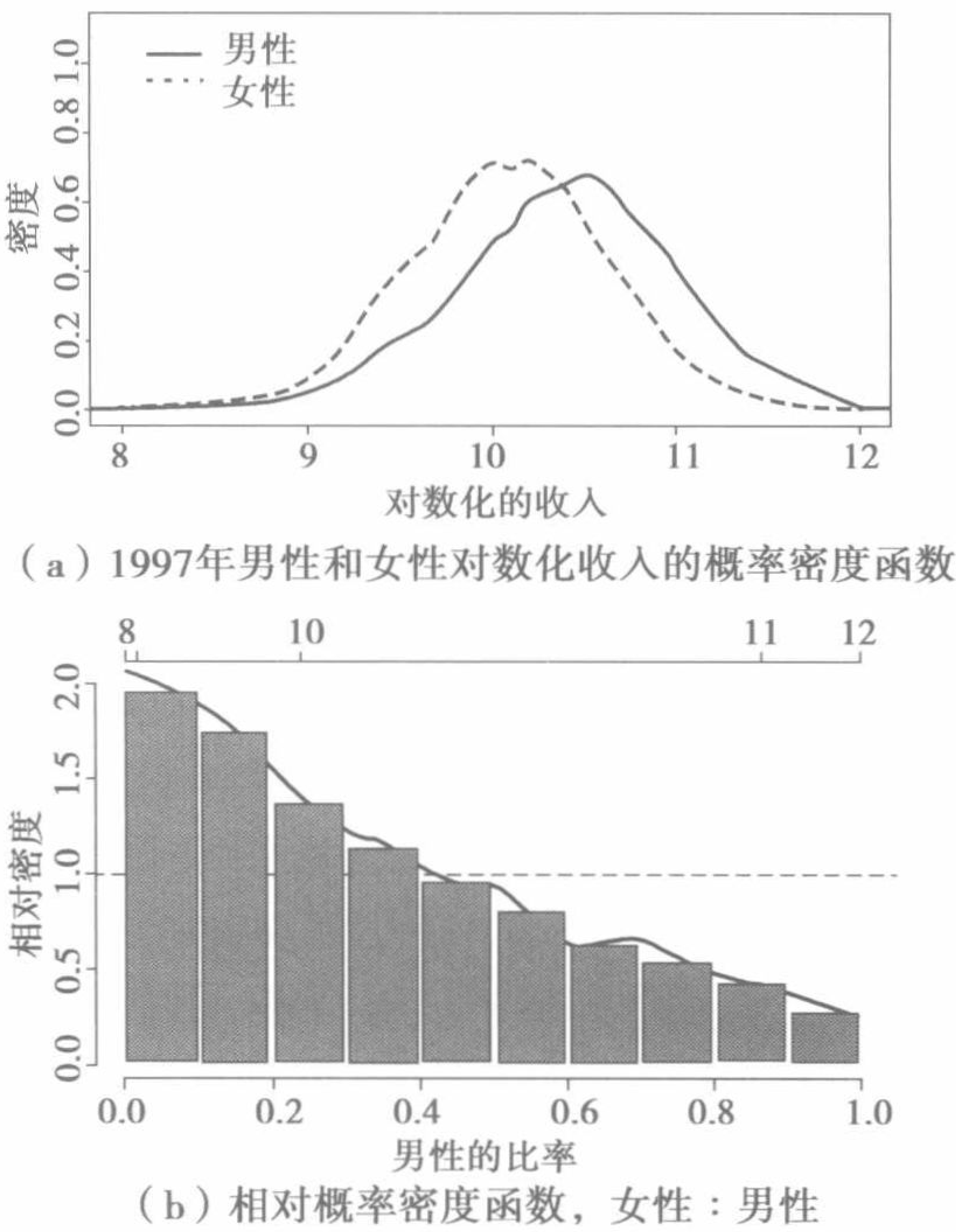
一个变量分布上的组间差异或历时的变化是社会科学研究的共同关注点。传统的参数模型仅限于对条件均值【Means】和方差【Variances】的分析,遗留了关于分布的大量信息未予处理。

相对分布法的目标在于,绕开以均值为基础的比较,对分布差异进行细致的分析。因此,它们给出了一个用以比较分布分析的一般框架。

相对分布给出了一个能够简化探索性数据分析的图形展示,一个统计有效的基础,该基础可用于发展统计假设驱动的综括量度、位置和形态,以及能够辨识组内和组间的分布变化来源的协变分解(covariate decompositions)。

相对分布法的主要思想可以通过其在收入比较中的一个应用来了解。图1(a)中的概率密度函数(probability density functions,PDF)显示了1997年全年全职工作的男性和女性的对数化的年收入。女性的分布明显偏左,但这个图形展示并没有给出任何额外的信息能用于比较。

相对分布(relative distribution)是一个分布中的观测值如果被放置到另一个分布中百分位等级的集。在这个例子中,它是:如果将女性收入者放置在男性收入的分布上百分位等级的集。



(a)1997年男性和女性对数化收入的概率密度函数

(b)相对概率密度函数,女性:男性

图1

图1(b)显示了女性收入相对于男性收入的相对分布密度。图中的平滑线代表在收入尺度的每个水平上,女性对男性的相对频、相应的比值显示在纵坐标上;顶部的横轴显示了对数化收入的美元值,底部的横轴显示了男性分布的百分位等级。直方图代表了女性落在男性收入分布的每一个十分位区间内的部分。我们可以从直方图上看出,20%的女性落在男性收入分布的最底部的十分位区间内,同时,18%的女性位于第二个十分位区间内。总的来说,75%的女性赚的少于处于中点的男性(前5个十分位区间的总和)。相比而言,不到5%的女性收入达到了男性收入分布的最高一个十分位数。

男性和女性分布的差异可以分成两个基本构成:位置的差异和形态的差异。如果女性的收入分布仅是男性分布的简单下移,那么,在匹配了中位数(或其他位置参数)后,两个分布应当相同。在进行了位置校准后,还遗留的差异即分布形态上的差异。

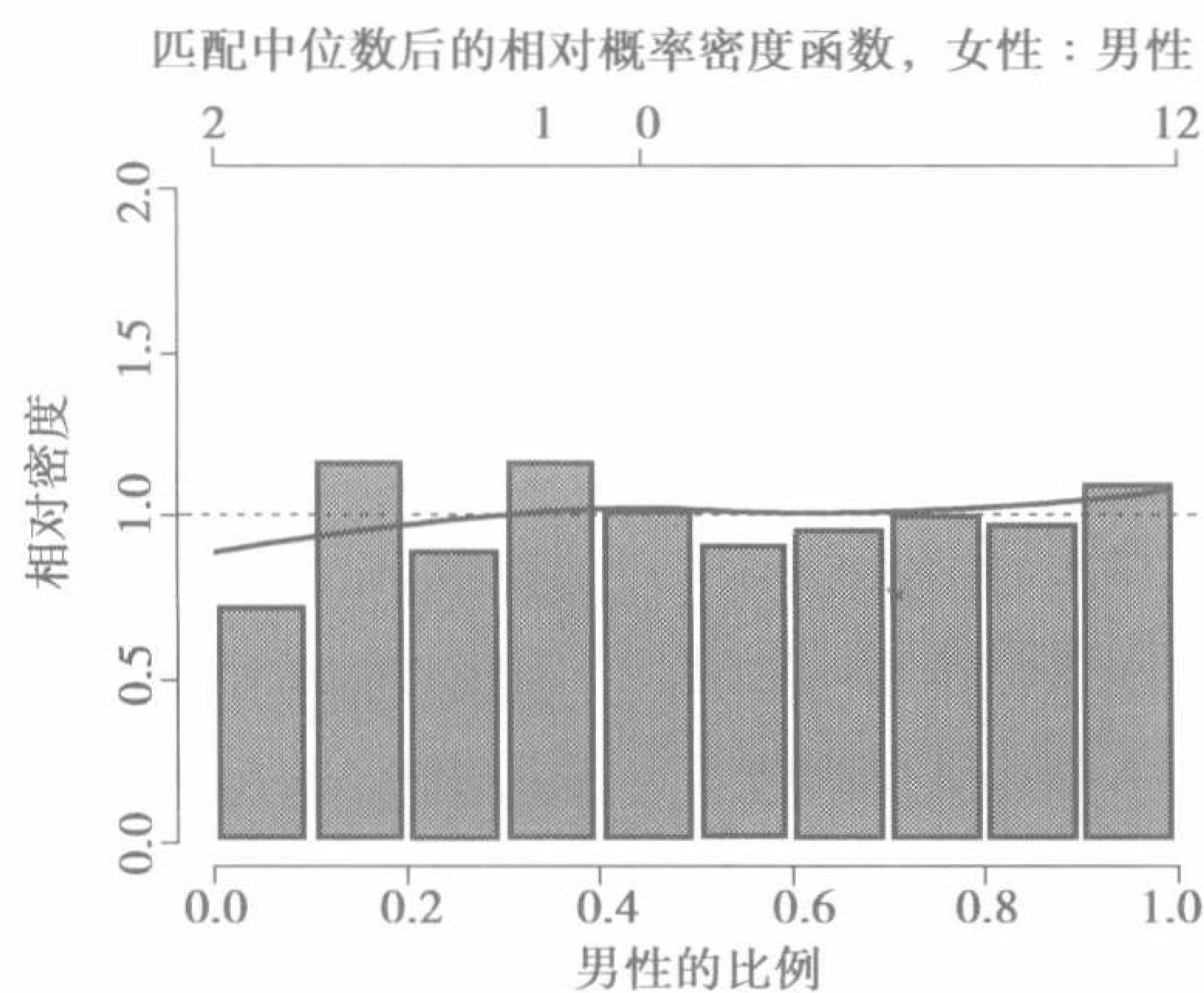


图 2

相对分布能被分解为这些位置和形态的成分,图 2 显示了男性和女性收入残余的形态差异。通过将女性收入分布对男性的收入进行中位数匹配(中位数收入比值大约为 1.4),进而构建新的相对分布,可以构建残余的形态差异图。进行中位数差异的校准后,新的相对分布几乎是平的,而这表示,两个组间的大多数差异可归之于女性收入的中位数下移。然而,仍存在着残余差异。

在低尾处,女性收入有相对小的密度(因为最低的十分位区间内比值于 1),而在高尾处,有较高的密度(图 2)。这或许反映了在过去 20 年中,低收入男性的急剧减少,以及高收入女性的相应增加。低尾部分的差异也可能标志着最低工资的效应,因为,最初女性的中位收入较低,同男性的情况相比,她们的最低收入则更接近于她们的中位收入。关于这种趋势的更多内容可以在莫里斯和韦斯顿(Morris & Western,1999)的评论文章中找到。汉德科克和莫里斯(Handcock & Morris,1999)以一本专著的篇幅处理了相对分布的理论,并囊括了相应的历史文献。

——Mark S. Handcock

Martina Morris

(马妍译　赵锋校)

参考文献

Handcock, M. S., & Morris, M. (1999). *Relative distribution methods in the social sciences*. New York: Springer-Verlag.

Morris, M., & Western, B. (1999). Inequality in earnings at the close of the 20th century. *Annual Review of Sociology*, *25*, 623-657.

相对频次(Relative Frequency)

运用统计方法的两个主要原因之一是综括和描述数据【Data】(另一个原因是用数据进行统计推论【Inference】)。相对频次是通过计算落入某个区间(适用于定量变量【Quantitative Variable】)或者属于某个类别(适用于分类【Categorical】变量)的观测值的比率,来综括复杂数据的一个简单工具。

某个区间(或某个类别)的相对频次等

于这个区间(或类别)内的观测值数量除以观测值的总数。一系列区间或类别,以及它们的相对频次被称为相对频次分布。如果将数据归入不同类别或区间的方式是恰当的,那么,呈现一个变量的相对频次分布可以给出该数据的分布的一个完整面貌。表1是一个定量变量——智商的相对频次分布示例。表2是一个定性变量【Qualitative Variable】——职业地位的相对频次分布示例。

表1 智商的相对频次(以5个单位为分组间隔)

智商	频次	相对频次	百分比
81~85	1	0.01	1.0
86~90	1	0.01	1.0
91~95	9	0.09	9.0
96~100	14	0.14	14.0
101~105	12	0.12	12.0
106~110	11	0.11	11.0
111~115	39	0.39	39.0
116~120	7	0.07	7.0
121~125	1	0.01	1.0
126~130	1	0.01	1.0
131~135	3	0.03	3.0
136~140	1	0.01	1.0
总计	100	1.00	100.0

表2 职业地位的相对频次

职业	频次	相对频次	百分比
专业人员	33	0.33	33.0
蓝领	21	0.21	21.0
手工艺者	25	0.25	25.0
白领	12	0.12	12.0
佣工	9	0.09	9.0
合计	100	1.00	100.0

表1中第一个区间的相对频次是1/100=0.01。此即,就相对频次为0.01而言,在100个人中有1个人为81~85分。注意到这个假想分布是以5个单位的互斥间隔来安排数据的。同样,表2中第一个类别的相对频次是33/100=0.33。可以将其解释为“0.33的相对频次,意味着100个人中有33个人从事专业人员的工作”。

数据分析者通常会以报告某个间距或者类别的百分比来代替相对频次。某个间距或者类别的百分比就是其相对频次乘以100。表1和表2中最右边一列报告了每个类别和分组间距的百分比。因为相对频次是比值,所以相对频次的总和应等于1。很显然百分比的总和应当等于100。然而,有时取整过程可能导致稍有出入的总和,如0.998(相对频次)或者100.01(百分比)。

注意,使用相对频次或百分比意味着数据具有相当的稳定性。因此,如果个案的总数非常小,比如15个,那么必须谨慎地使用相对频次分布。布莱洛克(Blalock,1979)提出了两条经验法则:一是,在报告相对频次或百分比的同时也报告观测值数量(如频次);二是,除非总的观测值接近或超过50,否则不要报告百分比或者相对频次。

同样值得注意的是定量数据分组的间距宽度会对相对频次分布的外形有重要的影响。例如,将表1中的数据以25个单位为间距重新分组,会根本改变相对频次的形态,与表1相比,给出的信息也更少。

然而,对“我是应当使用一个较窄的间距以得到更多的细节,还是应该使用一个大一些的间距以得到更加浓缩的数据”这样一个问题并没有固定的答案。它通常取决于数据的性质和研究的目的。一个惯例,即10~20的间距数可以在浓缩和细节间获得很好的平衡。在海斯(Hays,1994)及约翰逊和巴塔查

理亚(Johnson & Bhattacharyya,2001)的书中可以得到更多报告相对频次的信息。

最后,无论数据的测量层次是什么,数据分析者都能进一步使用直方图【Histogram】,以图形的方式呈现所有类别频次分布。直方图及其变化的样式,如条形图【Bar Graphs】和饼图【Pie Charts】都是刻画相对频次的有效方式。关于创建直方图的一般原则的讨论详见克利夫兰(Cleveland,1993)。

——Cheng-Lung Wang

(赵锋译校)

参考文献

Blalock, H. M. (1979). *Social statistics* (rev. 2nd ed.). New York: McGraw-Hill.

Cleveland, W. S. (1993). *Visualizing data*. Summit, NJ: Hobart.

Hays, W. L. (1994). *Statistics* (5th ed.). Fort Worth, TX: Harcourt.

Johnson, R. A., & Bhattacharyya, G. K. (2001). *Statistics: Principles and methods* (4th ed.). New York: Wiley.

相对频次分布(Relative Frequency Distribution)

一个相对频次【Relative Frequency】分布是几种用于描述一个相对较大数据集的图之一。图1和图2给出了相应的例子。图1是以英寸为单位测量的20~25岁男性的身高变量。图2是常用于美国教育系统,以字母A~F划分的课程评定等级。这些图被称为"相对"频次分布,因为y轴是一个百分比量尺,而不是一个频次量尺。如果200个学生中有20个得A等,那么这一组的相对频次是0.10。

要制作这些图,首先需将整个数据集分成不同的间隔或组别。间隔或组别画在图的x轴上,同时标出间隔中点或组别标签。条形(或柱形)在x轴上的垂直方向延伸到y轴代表间隔或组别相对频次的点。

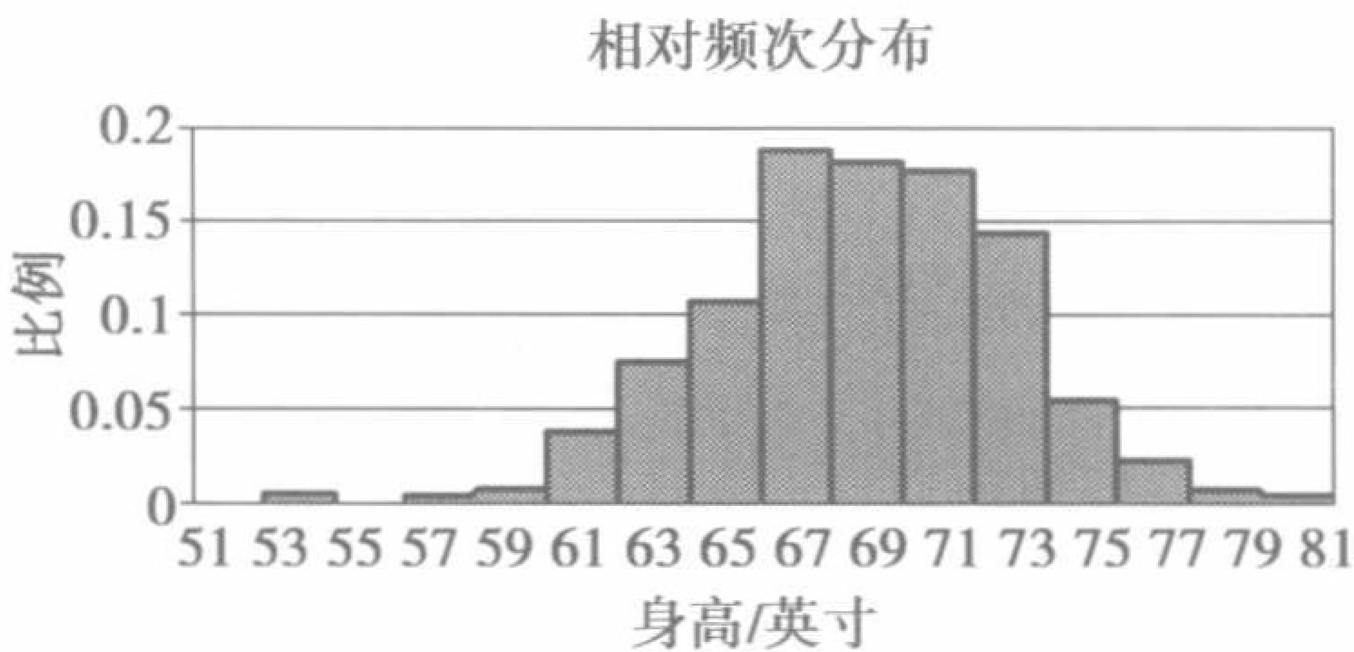


图1 男性身高分布的相对频次分布(从52~82英寸,以2英寸为间隔)

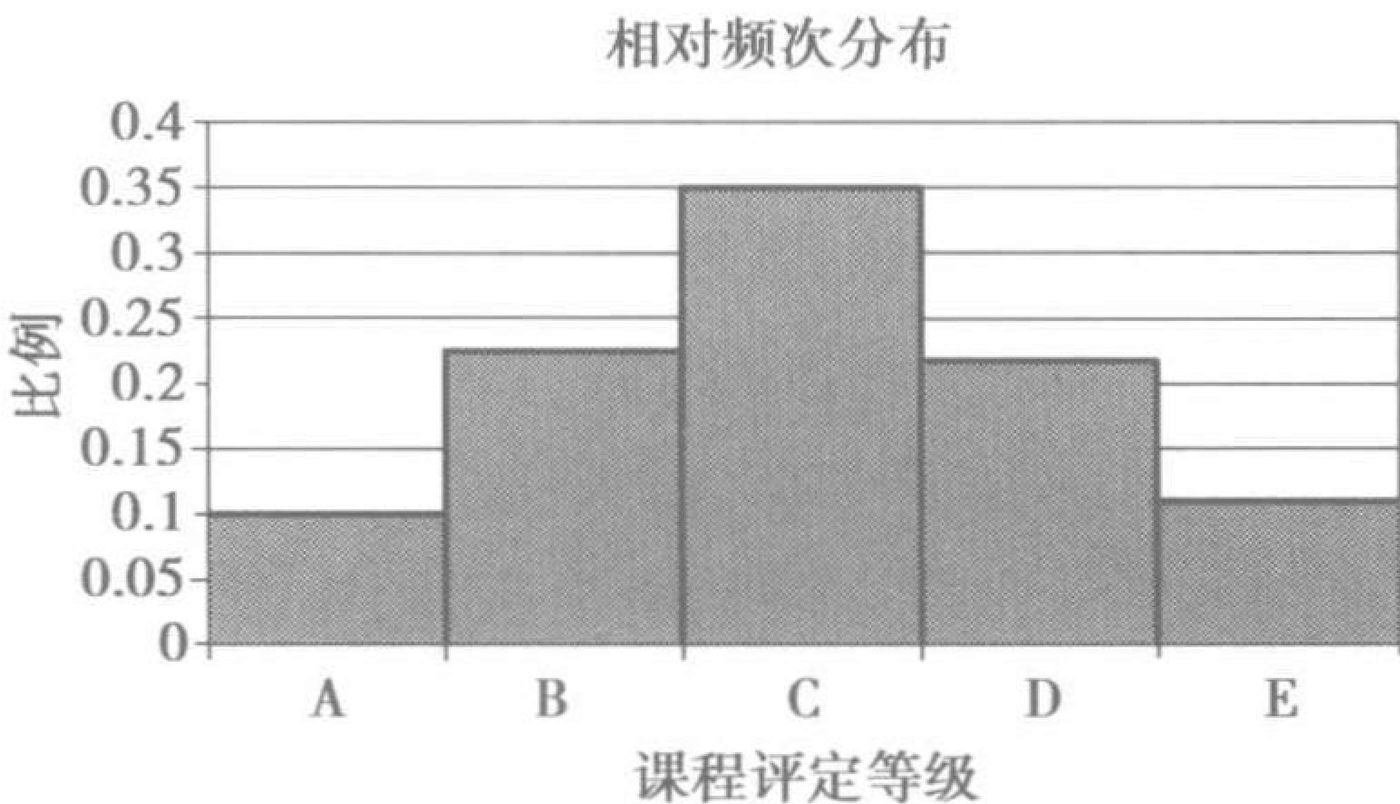


图2 课程评定等级的相对频次分布

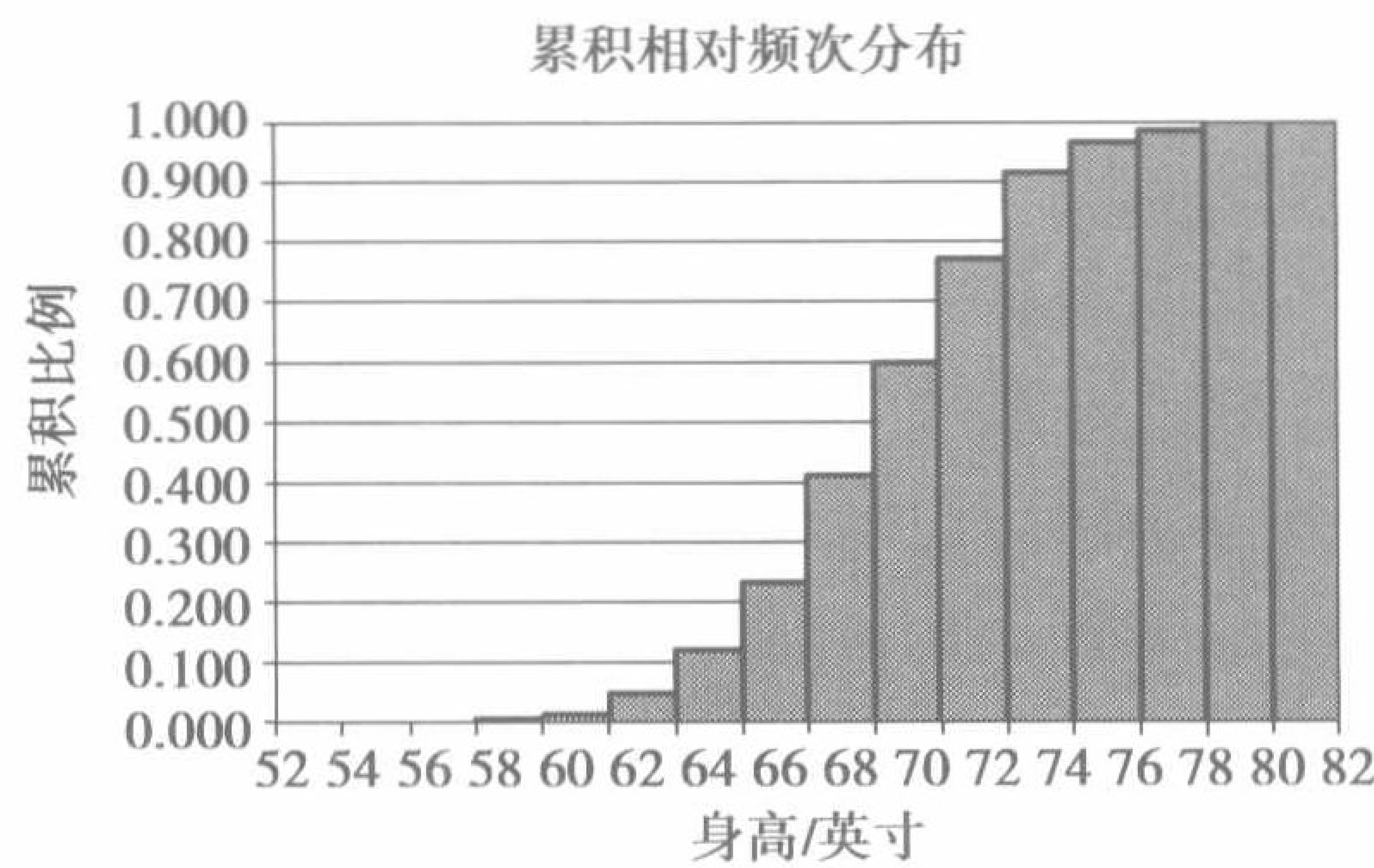


图 3 图 1 中身高数据的累积分布；
y 轴取值×100 即为百分比

相对频次分布的一种变化样式是累积比例图。使用这种方法时，将比例乘以 100 即为百分比。图 3 是身高的相对频次数据的一个累积频次图。要注意的是，x 轴的取值是间隔的上限，而不是之前间隔的中点。

——Kenneth O. McGraw

（马妍译 赵锋校）

相对变异(度)(Relative Variation [Measure of])

相对变异系数(coefficient of relative variation, CRV)是一个变量的标准差【Standard Deviation】除以其均值【Mean】所得到的标准化。这个量度常用于反映标准差的大小。耶尔和肯德尔(Yule & Kendall, 1968, pp. 143-144)将其视为分散度的一个绝对量度，因为用均值除控制了变量的测量单位。

相对变异系数的计算公式是 $CRV = s/\overline{X}$，其中 s 是变量的标准差，$\overline{X}$ 是变量的均值(参见变异系数【Coefficient of Variation】)。这个统计量只适用于定比变量，因为一个定距变量的量尺的变换会改变其均值，进而改变 CRV。除以临近 0 的值会得到一个很大的值，所以当均值非常接近于 0 时，相对变异系数被认为是不可靠的。公式的一种变化样式是用均值的绝对值作分母，这样 CRV 的值总是正的。第 i 个观测值的相对变异度可定义为：第 i 个值 x_i 与其均值的偏差【Deviation】除以均值，即 $RV_i = (X_i - \overline{X})/\overline{X}$。那么相对变异系数等于平方的观测值的相对变异度之平均数的开方。

$$CRV = \sqrt{\frac{\sum (RV_i)^2}{N}}$$

在比较均值不同的变量的差异度时，相对变异系数很有效。例如，试想有一组城市具有非常不同的平均犯罪数量，同时欲比较各城市一年中各月份犯罪率的方差。仅仅因为纽约的犯罪率高很多，纽约市各月份间的犯罪率方差就一定要比皮奥利亚市大很多。然而，用它们各自的均值去除各自的标准差会得到一个标准化的结果，而这个结果会更好地反映出纽约市一年中各月份间的犯罪率是否比皮奥利亚市更具可变性。

表 1　假想的年龄和教育数据

人	年龄	受教育年限
1	20	15
2	30	21
3	40	18
4	50	12
5	60	9
6	70	6
7	80	3
均值	50	12
标准差	20	6
相对变异系数	0.40	0.50

这个系数对于比较以不同单位测量的变量的差异度也很有效。例如,人们的年龄是不是比受教育年限有更大的变异度?成年人年龄的方差确实大于他们受教育年限的方差,但是,当用标准差除以均值进行标准化后,情况可能就不同了。比如,有一组成年人,他们的年龄均值是 50,标准差是 20,年龄的相对变异度是 0.4。如果他们平均受教育年限是 12 年,标准差是 6,那么受教育年限的相对变异度是 0.5(表 1)。在这个例子中,年龄的方差大于受教育年限的方差,但是一旦将均值纳入计算,教育实际上有更大的相对变异度。

——Herbert F. Weisberg

(马妍译　赵锋校)

参考文献

Blalock, H. M. (1979). *Social statistics* (rev. 2nd ed.). New York: McGraw-Hill.

Snedecor, G.W. (1956). *Statistical methods*. Ames: Iowa State College Press.

Weisberg, H. F. (1992). *Central tendency and variability*. Newbury Park, CA: Sage.

Yule, G. U., & Kendall, M. G. (1968). *An introduction to the theory of statistics*. New York: Hafner.

相对主义(Relativism)

相对主义是对客观主义【Objectivist】、本质主义【Essentialist】、基础立场的系统一种彻底地怀疑的姿态。它不太像某种反立场(antiposition)或某种元立场(metaposition)。相对主义者的论点强调在将科学主张做成客观的和基本的过程中,修辞学【Rhetoric】具有不可或缺的作用,同时,强调那些被宣称能保证客观基础的科学事物的偶然性,如检验、准则、规则、实验,以及其他程序的偶然性。相对主义者不否认判断的可能性(如其经常宣称的),但是,强调判断与更广泛的人类价值、文化、生活方式及政治观联系在一起,不能简化为客观的标准。在社会科学领域,相对主义与更广泛的理论发展相联系;它可以被视为后现代主义【Postmodernism】和建构主义,以及言语转向的哲学的相关物。

在社会科学领域,相对主义曾是最具争议,也最常被鄙视的立场之一。说一种分析的方法或形式"滑向相对主义",常被认为是足以摈除它的意思。作为一个相对主义者曾被认为是等同于支持(或至少不批评)大屠杀或虐待配偶,即站在一种认为不可能作出判断和"任何事情均可发生"的立场上。

近来,关于相对主义的争论常集中于唯实论【Realism】和政治的或伦理的批评之间的关联上。有些实在论者论说,相对主义不具备充分的手段,以发展政治主张。另一些人则鼓动起"死亡和家具"的辩论,即或者强调熟悉的和惊人的暴行(1991 年在巴士

拉路上对正在逃跑的伊拉克士兵的屠杀,女性的割礼)或者突然拍打桌面(Edwards, Ashmore, & Potter, 1995)。类似的争论也出现在女性主义中,其中不同的观点,就性别、真理的性质及政治诸论题,常发生争议(Hepburn, 2000)。

罗蒂(Rorty,1991)沿着更倾向于相对主义的路线,就认识论、伦理学和政治哲学中的问题进行了再思考。他借用雅克·德里达的解构论哲学(deconstructive philosophy)来说明写作在构建某种观点、下定判断,以及构建立场,使其显得坚实且永恒的过程中所起的作用。他的反基础主义(antifoundationalism)的论点认为,判断被植入断言的方式本身在社会性上是偶然的。芭芭拉·赫恩斯坦·史密斯(Smith,1997)论说,借助于对真理、理性和其他客观主义的核心观念的失效的相对主义的确认,政治学、道德体系及文化评价将变得更充实,而不会被损害。

哈利·柯林斯主张,对科学知识的研究如果要具有逻辑上的一致性,方法论上的相对主义就是必要的。它是一场新的对科学的社会学研究运动之主力。这一主张的要点在于,不把任何领域中流行的科学正统视作研究起点,无论这些科学正统看起来多么有道理。因为,只有这样做才能不再进行非对称的分析(即把科学正统看作得到科学观测保证的知识,替代做法则要求对正统作出社会解释)。例如,在研究引力辐射的科学理论时,柯林斯研究了修辞术和说服术扮演的角色。正是通过两种技术的运用,某些实验被做成可以重复实现的样子,一些断言则人为地被做成不可信的样子。柯林斯试着将主张和反对重力波理论的研究者放在同一天平上考察,而不是根据传统的观点,在开始时,就认为研究者已经发现重力波并不存在。就此类分析而言,研究者不认为,科学争论可以由一组中立于社会之外的规则和准则来决断,反而是把重复、可检验(testability)的这类准则变成争论的属性。

相对主义在科学知识的社会学中的核心地位是它在当代社会科学领域占据巨大重要性的原因之一。在此基础上,马尔科姆·阿什莫尔(Ashmore,1989)发展了反身性的立场,反复要求一种比柯林斯的立场更加彻底的相对主义。在科学知识社会学中,方法论相对主义现在已经被广泛接受。乔纳森(Potter,1996)论说,在分析科学领域之外的"对事实的说明"时,需要类似的考量;否则,偏袒一方,强化认知现状的同一问题会再次出现。

关于相对主义思考的道德和政治后果的热议还在持续。

——Jonathan Potter

(王玥译　赵锋校)

参考文献

Ashmore, M. (1989). *The reflexive thesis: Wrighting sociology of scientific knowledge*. Chicago: University of Chicago Press.

Collins, H. M. (1985). *Changing order: Replication and induction in scientific practice*. London: Sage.

Edwards, D., Ashmore, M., & Potter, J. (1995). Death and furniture: The rhetoric, politics, and theology of bottom line arguments against relativism. *History of the Human Sciences*, *8*, 25-49.

Hepburn, A. (2000). On the alleged incompatibility between feminism and relativism. *Feminism and Psychology*, *10*, 91-106.

Potter, J. (1996). *Representing reality: Discourse, rhetoric and social construction*. London: Sage.

Rorty, R. (1991). *Objectivity, relativism, and truth: Philosophical papers volume 1*. Cambridge, UK: Cambridge University Press.

Smith, B. H. (1997). *Belief and resistance: Dynamics of contemporary intellectual controversy*. Cambridge, MA: Harvard University Press.

信度(Reliability)

自查尔斯·斯皮尔曼在1904年奠定了该理论的基础后,信度理论主要在心理学相关学科领域首先得到了发展。然而,这些领域对信度的强调并不意味着这个问题与其他学科无关。信度能告诉研究者,个体对某一测量(measure)的反应的随机不一致性(random inconsistency)或者非系统波动(unsystematic fluctuation)的相对量。一个量表的信度是一个量表的构念效度【Construct Validity】获得支持的诸证据中一项必要的证据,虽然一个可靠的量表并不必然是一个有效的量表。因此,任何量表的信度都为构念效度设定了一个上限。

个体对一个量表的反应倾向于随环境的不同而非系统的变化。这些非系统的变化被称为随机误差【Random Error】或随机的测量【Measurement】误差,会造成回答人不一致的反应。测量的随机误差源于多种因素,而这些因素既不可预知,也不具有规律的模式。这些因素包括测验接受者的心理状态和生理状况、测验环境的特征、测验实施的不同场合、测试实施者的特征或测验题项(其中某些题项措辞模糊或措辞不当)的不同样本。由此以来,一个结果是,如果个体被多次测量,上述因素之一或其中的多个因素的组合就会导致个人的反应出现不可预知的波动。

测量误差的类型

可以说绝大部分(如果不是全部)的量表都是不完美的。根据测量理论最近的发展,任何被测变量的分值可能由三个可区别的成分构成:目标概念、非目标概念、随机的测量误差(Judd & McClelland, 1998)。其中,后面两个成分都属于测量误差,它们虽然不是所想要测量的内容,但是在实践中却无法排除。测量分中代表非目标概念的部分被视为系统误差【Systematic Error】。在回答人的分值上,这类误差会产生规整的但无关的变化。相反,一个量表的随机测量误差不会将任何规律性模式加于回答人的分值上。应当注意的是,对个体而言,一个分值的测量误差并不一定是一个完全随机的事件(比如,某学生的测验焦虑);不过,如果面对的是非常多的回答人的得分,测量误差的成因可能会随机地变化,其行为好似随机变量【Random Variable】。因此,可将这类测量误差假定为随机的。信度理论试图理解、量化和控制的正是随机的测量误差。

一个量表上随机误差的效应

对量表的非一致反应(inconsistent responses)对科学质询和实践应用具有重要含义。吉赛利、坎贝尔和泽德克(Ghiselli, Campbell, & Zedeck, 1981)指出,对某个量表的非一致反应无甚价值,无论研究者的计划是以同一量表为基础比较两个或多个个体,或根据受试人的能力将其归于某一组中,或预测一个人的行为,或评估某一干涉的效应,还是发展相关理论。试想一下,一个大学的招生官员想根据一次入学考试,从两名入学申请者中选出一名合适的入学者。如果就此项测验,申请人的回答从这一次到下一次之间倾向于有20分的波动,那么两个申请人之间在一次测验中实际发生的10分差值就可能是由随机测量误差带来的。于是,仅根据两者间的10分差值所作出的推论,即在大学期间,一个申请人比另一个

更有能力取得成功,可能就是错误的。同样,如果一个人在接受了发怒管理治疗后,在某一发怒量表上的得分减少了 10 分,这也可能源于对量表的随机反应。源自一种或多种因素的随机测量误差的量可以用各种信度系数予以测定,而这正是下一部分要讨论的。

经典的真实分理论

自 1904 年以来,发展出两种主要的信度理论:以平行测验模型为基础的经典真实分理论和以范畴抽样(domain sampling)模型为基础的概化理论【Generalizability Theory】。就历史发展而言,斯皮尔曼(Spearman,1904,1910)首先开始发展经典真实分【True Score】理论,而后瑟斯通(Thurstone,1931)、格利克森(Gulliksen,1950)和吉尔福德(Guilford,1954)推进了该理论的发展。概化理论是一个新理论,由杰克逊(Jackson,1939)和特赖恩(Tryon,1957)提出,而后由洛德和诺维克(Lord & Novick,1969)以及克龙巴赫和他的同事(Cronbach, et al.,1972)将其推向深入。传统的信度系数,如平行复本(parallel-form)信度和备择复本(alternate-form)信度、再测信度【Test-Retest Reliability】、内在一致性信度、折半信度【Split-Half Reliability】、评分者信度【Interrater Reliability】,以及编码者内(intra-coder)信度,都是以经典的真实分理论为基础的信度估计量,尽管这些系数也可以用概化理论(详见 Judd & McClelland,1998)进行估计。由于概化理论及其应用在百科全书的其他部分有叙述,在余下的部分仅集中讨论经典的真实分理论。

经典的真实分理论是一个描述和量化随机测量误差的简单、有效且广为应用的理论。一般而言,经典理论的推理基于如下假定:

1. $X_i = T + E_i$
2. $E(X_i) = T$
3. $F(X_1) = F(X_2) = F(X_3) = \cdots = F(X_k)$
4. $\rho_{X_1X_2} = \rho_{X_1X_3} = \rho_{X_1X_k} = \cdots = \rho_{X_{(k-1)}X_k}$
5. $\rho_{X_1Z} = \rho_{X_2Z} = \rho_{X_3Z} = \cdots = \rho_{X_kZ}$
6. $\rho_{E_1E_2} = \rho_{E_1E_3} = \rho_{E_1E_k} = \cdots = \rho_{E_{(k-1)}E_k} = 0$
7. $\rho_{E_1T} = \rho_{E_2T} = \cdots = \rho_{E_kT} = 0$

假定 1,$X_i = T+E_i$,宣称,一次测验 $i(X_i)$ 中的一个测得分数(observed score)(或者易错分数,fallible score)由两部分组成,即真实分数(T)和随机的测量误差(E_i)。由前文提及的各种因素的组合引起的随机测量误差在不同的平行复本测验中或者是正值或者是负值,因此,这就使得测得的分数或者高于真实分数,或者低于真实分数。

一项测验中的一个真实分数并不代表一个人的某种特质的真实水平。换言之,由于它来自诸多因素的一个组合,因此它代表的是对该测验的一致的(consistent)反应。注意,诸多因素的不同组合会引出多种多样的一致的反应。例如,在潮湿、嘈杂和经常被打断的环境中,对一个智力量表的一致的反应完全不同于在凉爽、安静和可控环境中,对同一量表的一致的反应。前一种一致的回答代表艰难环境中的智力的真实分数,而后一反应是愉悦条件下的智力的真实分数。

假定 2,$E(X_i) = T$,宣称,如果对个体进行无限次平行的测验,真实分数是该人所有测得分数的期望值(均值)。例如,假设某个人接受了 k 次(很多次)平行复本测验,这个人每次测验的测得分数可以表示如下:

$$
\begin{aligned}
X_1 &= T + E_1 \\
X_2 &= T + E_2 \\
&\vdots \\
X_k &= T + E_k
\end{aligned}
$$

因为测量误差或者是正的或者是负的，所以在一个很长的序列中，这些误差的和为0，即测量误差的期望，$E(E_i) = 0$。因此所有测得分数的期望值等于真实分数。

假定3到假定5涉及平行复本测验的特征，而这些假定既有理论意义，也有实践意义。假定3，$F(X_1) = F(X_2) = F(X_3) = \cdots = F(X_k)$，宣称，就一群数目众多的应试者而言，他们在每一次平行复本的测验中测得分数的分布都是同一的。根据同分布（identical distributions）假定，我们期望，$T_1 = T_2 = \cdots = T_k$，并且 $\sigma^2_{X_1} = \sigma^2_{X_2} = \cdots = \sigma^2_{X_k}$。此外，$k$ 次平行复本测验的测得分数之间的相关系数相同，即表示为 $\rho_{X_1X_2} = \rho_{X_1X_3} = \rho_{X_1X_k} = \cdots = \rho_{X_{(k-1)}X_k}$（假定4）。最后，假定5，$\rho_{X_1Z} = \rho_{X_2Z} = \rho_{X_3Z} = \cdots = \rho_{X_kZ}$，宣称，在 k 次平行复本测验上的测得分数与任意其他量表 Z（不同于这 k 次测验的量表）之间，有相等的相关系数。

假定6和假定7实质上宣称，测量误差是完全随机的。因此，一个大的应试者总体【Population】在各 k 次测验上的测量误差之间，以及测量误差和真实分数之间，其相关系数都等于零。基于假定7，σ^2_X，σ^2_T 和 σ^2_E 之间的方差【Variance】关系能够进一步表示为 $\sigma^2_X = \sigma^2_T + \sigma^2_E$。换句话说，在一个应试者总体中，测得分数的方差等于真实分数的方差加测量误差的方差。将上述方差关系同假定3结合起来，可推知，k 次平行复本测验的测量误差的方差彼此相等，$\sigma^2_{E_1} = \sigma^2_{E_2} = \cdots = \sigma^2_{E_k}$。

信度的理论定义

根据方差关系 $\sigma^2_X = \sigma^2_T + \sigma^2_E$，当随机误差不存在（如 $\sigma^2_E = 0$ 或 $\sigma^2_X = \sigma^2_T$）时，会出现一个完全可信的量表。如果增大，一个量表上观测到的差值将同时反映回答人之间的真实分数的差值和随机误差的差值。当 $\sigma^2_X = \sigma^2_E$ 时，所有观测到的差值反映的只是随机的测量误差，这表明对这一量表的反应是完全随机的和不可信的。

如果在上述等式的两边同时除以 σ^2_X，方差关系即变为：

$$1 = \frac{\sigma^2_T}{\sigma^2_X} + \frac{\sigma^2_E}{\sigma^2_X}$$

一个量表的信度可定义为：

$$\frac{\sigma^2_T}{\sigma^2_X}$$

可以将它解释为测得分数和真实分数之间的相关系数的平方（ρ^2_{XT}）。在概念上，信度可以解释为测得分数的方差可以由真实分数方差解释掉或可归之于真实分数方差的比例。依据上述假定可进一步证明，两个平行复本测验之间的相关系数等于$\frac{\sigma^2_T}{\sigma^2_X}$。此即，

$$\rho_{X_1X_2} = \frac{\sigma^2_{T_1}}{\sigma^2_{X_1}} = \frac{\sigma^2_{T_2}}{\sigma^2_{X_2}}$$

如果两个平行复本测验之间的相关系数是0.80，那么它表明测得分数的方差的80%可以归于真实分数的方差。下面，将使用通行记号 $\rho_{XX'}$ 和 $r_{XX'}$，分别表示总体和样本【Sample】的信度。

估计信度的方法

各种信度系数都是以经典的真实分理论为基础的，如再测信度、内在一致性信度、评分者间信度或编码者内信度，而这些信度系数在百科全书的其他部分有介绍。我们在此只集中讨论另外三种信度系数：备择复本信度、差值分（difference scores）信度和复

合分(composite scores)信度。

备择复本信度系数,也称为等价系数,其估计是通过计算同一量表的两个相似的测量形式(具有相似的均值、标准差【Standard Deviations】以及与其他变量的相关系数)所得到的测得分数间的相关系数。通常,一组回答人会在一次测验中,在一段较短的时间内以随机顺序接受两种形式的测量,尽管两次测验实施间的时间间隔可能稍长。

在进行干预或评估心理和生理的发展时,研究者通常的兴趣在于前测和后测得分上的变化。同一量表或两个可比量表在两个不同时点上测得分数的差值常被称作一个变化分或差值分(记作 D)。一个样本的差值分信度可用如下方程估计:

$$r_{DD'} = \frac{s_X^2 r_{XX'} + s_Y^2 r_{YY'} - 2 s_X s_Y r_{XY}}{s_X^2 + s_Y^2 - 2 s_X s_Y r_{XY}}$$

其中,s_X^2 和 s_Y^2 是前测和后测的方差,$r_{XX'}$ 和 $r_{YY'}$ 是测验的信度,而 r_{XY} 是两次测验之间的相关系数。如果两个量表的方差相等,上述等式可改写为:

$$r_{DD'} = \frac{\frac{r_{XX'} + r_{YY'}}{2} - r_{XY}}{1 - r_{XY}}$$

假定 $r_{XX'}$ 和 $r_{YY'}$ 是常量,那么较强的 r_{XY} 会导致 $r_{DD'}$ 较弱。此外,当 r_{XY} 趋近于两个量表的平均信度时,$r_{DD'}$ 趋向于0,实际上,因为两个量表是类似的,所以 r_{XY} 的强度通常较强。其结果是:即使两个量表各自相对可信,差值分也倾向于有低的信度。

对应于差值分,研究者有时需要使用一个复合分——即 k 个量表的原始得分的一个线性组合。例如,一位人事心理学家可能会根据一项智力测验、一项心理运动能力测验、一份兴趣清单,以及一个个性量表的一个复合分来选取工作申请人。这一复合分(C)的信度可以估计如下:

$$r_{CC'} = 1 - \frac{\sum_{i=1}^{k} s_i^2 - \sum_{i=1}^{k} (r_{ii'} s_i^2)}{s_C^2}$$

其中,s_C^2 和 s_i^2 是复合分的方差,而 $r_{ii'}$ 是第 i 个量表的信度。在实践中,研究者可能用 k 个量表的标准分【Standardized Scores】的一个线性组合作为复合分。这个复合分的信度可用简化的公式进行估计:

$$r_{CC'} = 1 - \frac{k - k\bar{r}_{XX'}}{k + (k^2 - k)\bar{r}_{XY}}$$

其中,$\bar{r}_{XX'}$ 是 k 个量表的平均信度,而 $\bar{r}_{XY}$ 是各测验之间平均的两两相关系数。假设测验间的两两相关系数是常量,复合分的信度就比各量表的平均信度大。因此,一个复合分的信度倾向于相对地更可靠。

测量的标准误

JP2 如前所述,信度系数只给出个体对某一量表的反应的随机不一致度的相对量。一项 Cronbach 系数为 0.90 的检验表明该检验比另一个 Cronbach 系数为 0.70 的检验具有更高的内在一致性。然而,两个信度估计系数都标明测验分并非绝对精确。例如,斯坦福-比纳(Stanford-Binet)智力量表是一个非常可靠的量表,具有 0.95~0.99 的内在一致性信度。虽然这一测验高度可信,但我们不能肯定地说60 的得分确实比平均得分 50 要高出 10 分。更实际地说,这 10 分反映的是,智力的真实差值加随机的测量误差。

为了准确地解释测验得分,研究者需要将不可信性(unreliability)纳入考虑,即通过测量误差方差的平方根或者测量的标准误(σ_E)来测定不可靠性。根据公式

$$1 = \frac{\sigma_T^2}{\sigma_X^2} + \frac{\sigma_E^2}{\sigma_X^2}$$

可得,一个总体中的测量标准误 $\sigma_E = \sigma_X\sqrt{1-\rho_{XX'}}$,而这一总体标准误 $\sigma_E = S_E = S_X\sqrt{1-r_{XX'}}$ 可予以估计。从概念上讲,测量标准误揭示了就此一测验而言,若给定测量误差的量级,那么可以期望一个测得分数有多大范围的变化。当 σ_E 已知时,研究者可用它来建构测得分数的置信区间【Confidence Intervals】。具体而言,一个样本中的测得分数(X)分别具有大约 68%、95%和 99%的信心落入 $X \pm 1\hat{\sigma}_E$、$X \pm 2\hat{\sigma}_E$ 和 $X \pm 3\hat{\sigma}_E$ 的置信区间内。

影响信度估计的因素

在估计一个量表的信度时,研究者难以分离回答人的特征和量表自身的特征。例如,一项代数测验的测验表现部分地取决于测验题目的难易。与此同时,测验题目难易的一个决定因素又有赖于测验接受者的能力(比如,大学生与初中生相比;同质群体与异质群体相比)。当一个群体较为同质时,比如说,其成员具有类似或相同的代数能力,测得分数的方差就会受到约束,即使测验本身完全可靠,这也会产生一个较低的信度。这个例子清晰地表明,基于经典的真实分理论的信度估计值会随测验接受者群体的情况及(或)测验题目的情况而发生变化。具体而言,估计信度的传统方法是受试者群组依赖的和测验内容依赖的,而这就暗示着,对于各种水平的回答人而言,一个量表的测量标准误(及信度)并不相同。实证证据显示,极端得分的测量标准误倾向于比普通得分的测量标准误小。

量表的其他特征也会影响信度。首先,根据斯皮尔曼-布朗预测公式(Spearman-Brown prophesy formula)假定一个量表已经包含了恰当的题项,那么随着该量表长度的增加,该量表的信度趋向于增加。斯皮尔曼-布朗预测公式界定如下:

$$\hat{\rho}_{XX'} = \frac{n \times \rho_{XX'}}{1 + (n-1)\rho_{XX'}}$$

其中,n 是一个原测验量表被加长的因子数,$\hat{\rho}_{XX'}$ 和 $\rho_{XX'}$ 分别是测验延长后和原测验的信度估计。其次,经验发现表明,若一量表的回答选项增加,信度系数也趋于增加。蒙特卡洛【Monte Carlo】研究揭示出:一个包含10个题项的量表的整个信度估计,如再测信度或 Cronbach α【Cronbach's Alpha】系数,会随着回答选项从一个 2 点量尺逐步增加到一个 5 点量尺而逐步增强。而后,信度估计变得稳定。测验的内容也会影响信度。例如,如果测验的时间不够充分,能力测验的信度可能被人为地放大。能力测验指的是那些用以检验回答人就某个主题的知识的测验。如果测验的时间很短,大多数回答人就不能在规定时间内答完所有问题。其结果是在各个测验题项上,他们的回答变得人为地一致。

最后,信度还会受到研究者所选择使用的估计方法的影响。如前文所述,基于经典的真实分理论的信度系数可以估计某一种或某两种测量误差。例如,内在一致性估计的是,那些由题项的异质引起的反应变化,或由测验内容的不同带来的反应变化;与之同时,再测信度估计的是,由于历时的变化,包括延时效应,所引起的反应变化。相应地,具有长时间间隔(如四个月)的备择复本信度估计的是两类随机误差,即那些由历时变化引起的和由测验内容的不同引起的。尽管所有基于经典真实分理论的信度系数估计的都是测量误差的相对量,但是用不同信度系数估计的误差在质上互不相同。因此,比较两种不同类型的信度估计是不符合

逻辑的。

当研究者纠正(correct)由于不可靠性产生的信度系数的衰减(attenuation)时,不同的随机误差需由不同的信度系数予以估计,这一事实具有重要意义。根据经典的真实分理论假定,两个真实分间的一个总体相关系数($\rho_{T_XT_Y}$)会受到量表(X 和 Y)的不可靠性的反向影响。该反向效应可根据量表 X 和量表 Y 的总体相关系数同各自总体信度($\rho_{XX'}$ 和 $\rho_{YY'}$)的关系推出,即界定如下:

$$\rho_{T_XT_Y} = \frac{\rho_{XY}}{\sqrt{\rho_{XX'}\rho_{YY'}}}$$

通常,研究者可能想指出的是,如果两个量表都完全可信,那么样本的相关系数 $\hat{r}_{XY}$ 会怎样。两个量表间未衰减的相关系数可由上述公式的修正公式得到,即

$$\hat{r}_{XY} = \frac{r_{XY}}{\sqrt{r_{XX'}r_{YY'}}}$$

当两个量表之一或两个量表变得更不可信时,$\hat{r}_{XY}$ 和 r_{XY} 之间的差会增加。例如,若 $r_{XX'} = r_{YY'} = 1$。然而,当 $r_{XX'} = r_{YY'} = 0.50$ 时,$\hat{r}_{XY}$ 的值会变为 r_{XY} 的 2 倍。在实践中,对于衰减的修正趋于高估两个量表间实际的相关度。首先,这一假设,即假设 $r_{XX'}$ 和 $r_{YY'}$ 处理的是同一类测量误差,是经不起推敲的。此外,存在着许多不同类型的,不适宜用经典的真实分理论予以估计的测量误差。例如,研究者的兴趣可能在于发现一项成就测验的可信度,而该测验是由男性实施者和女性实施者在两个不同时点以及三个不同的场所实施的。

——Peter Y. Chen

Autumn D. Krauss

(马妍译　赵锋校)

参考文献

Crocker, L., & Algina, J. (1986). *Introduction to classical & modern test theory*. New York: Harcourt Brace Jovanovich.

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 105-146). New York: Macmillan.

Ghiselli, E. E., Campbell, J. P., & Zedeck, S. (1981). *Measurement theory for the behavioral sciences*. San Francisco: W. H. Freeman.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

Judd, C. M., & McClelland, G. H. (1998). Measurement. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1). Boston: McGraw-Hill.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

定性研究的信度和效度(Reliability and Validity in Qualitative Research)

定性研究者【Qualitative Researchers】对信度【Reliability】和效度【Validity】的概念并没有单一或固定的定义。如同对定性质询的理解发展出不同的方向一样,特别是过去的 20 多年,对这两个概念的理解也是如此。

信度

在定量研究者那里,信度的原本表述是,如果一个研究可以被其他研究者重复开

展，那么这个研究被认为是可靠的。在定性研究者这里，对于下面的问题，即重复研究或再生成研究结果的可能性对判定定性研究的价值是否仍有相当的重要性，有着不同的思考。有些人说，例如基尔克和马克·米勒(Millar,1986)在观察的可重复性这个意义上，信度在定性质询中仍然扮演着重要的认识论角色。他们争辩，对一个被判定为好的或有效的研究而言，研究所做的观察必须是经久可靠的，同时，不同的方法(如访谈和观察)应当产生相似的结果。

其他人提出，特别是埃贡·古巴和伊冯娜·林肯(Guba & Lincoln,1994)，由于定性质询潜在的哲学假设(如现实是建构出来的)，信度的概念应由一个类似的可靠性(dependability)的观念所代替。他们争辩，对定性质询而言，重复是不可能的；对一个研究者所能期待的或应该期待的是，关于她或他是如何获得和分析资料的细致说明。

最后，当下有些人认为，继续把信度作为一个判断研究质量的标准是没有价值的。尽管一个研究者在某一情境下可能重复另一个研究者在那个情境中所做的观察和发现，但是，如果这种情况没发生，人们也不应该去担心。这种主张的基础源自以下认识，即理论-无涉的观察是不可能的，或者研究者只能从所在世界的某个特定位置观察世界(Smith & Deemer, 2000)。如果是这样的，那么信度就没有特殊的认识论的地位，也无须成为定性研究者的一个重要议题。

效度

对定量研究者来说，效度是一个主要的具有认识论价值的概念。说一个研究是有效的就意味着它是一个好的研究，即研究者在研究中准确地呈现了所考察的现象或现实。由于各种相关原因，特别是由于摈弃朴素的或直接的唯实论和认识论的基础主义(foundationism)，定性研究者发现这种效度是不可接受的。定性研究者提出了许多不同的立场，以取代实在论和基础主义，其范围从间接的或微妙的实在论及认识论的准基础主义(非确真论，fallibilism)，到其间伴随各种变化主题的非实在论和非基础主义(Smith, 1993)。于是，定性研究者努力去替换关于效度的传统理解，而这些努力导出了一系列相当令人困惑的定义和相应的变体。

马丁·哈默斯利(Hammersley,1990)是其中的一个代表。他基于一种微妙的或间接的实在论，并结合以准基础主义或非确真论的认识论，推进了对效度的理解。对他来说，效度的判定不在于一项说明能在多大程度精确地再现现实，而在于判定一项说明在多大程度上忠实于所考察的特殊情境。这意味着，即使我们没有直接接触到现实，并且我们的判断也总是可能出错，但也总是可以提出好的理由，从而能够宣称一项说明是真的。这些思索导出了关于一项有效研究的下述定义，即一项有效研究的结果经过似真性(plausibility)和可信性(credibility)的检验。似真性是指基于既有的关于所说明的情境的知识，对情境的说明是否可能为真。可信度考虑基于所研究现象的性质、研究的客观条件、研究者的特质等情况，研究者的判断是否准确。

第二个大方向，也在某种程度上，基于间接的实在论和准基础主义，涵盖了相当多样的、被称作“混编的”效度的定义。许多词，如催化的、情境的、撩人的、越界的、讽刺的及质问的，都作为修饰词加在效度上。大卫·阿什德和约翰·约翰逊(Altheide & Johnson,1994)以及帕蒂·拉瑟(Lather,1993)都讨论过上述词汇，以及其他的词汇。最后，有些定性研究者以为无须再为效度的

概念烦恼。因为这些人接受了非实在论和非基础主义的观点,所以他们也没兴趣继续讨论对世界的真实说明,以及对独立的既存现实的精确描述。他们声称,我们所能得到的一切都是以语言为中介,对现实的社会构建。于是,对于效度,最多可以说它是一项社会的一致同意,而此种一致同意不是认识论的,而是政治的和道德的(参见 Schwandt, 1996; Smith & Deemer, 2000)。

——John K. Smith

(王玥译 赵锋校)

参考文献

Altheide, D., & Johnson, J. (1994). Criteria for assessing interpretive validity in qualitative research. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (1st ed., pp. 485-499). Thousand Oaks, CA: Sage.

Guba, E., & Lincoln, Y. (1994). Competing paradigms in qualitative research. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (1st ed., pp. 105-117). Thousand Oaks, CA: Sage.

Hammersley, M. (1990). *Reading ethnographic research*. London: Longman.

Kirk, J., & Millar, M. (1986). *Reliability and validity in qualitative research*. Beverly Hills, CA: Sage.

Lather, P. (1993). Fertile obsession: Validity after poststructuralism. *Sociological Quarterly*, *35*, 673-694.

Schwandt, T. (1996). Farewell to criteriology. *Qualitative Inquiry*, 2, 58-72.

Smith, J. (1993). *After the demise of empiricism: The problem of judging social and educational inquiry*. Norwood, NJ: Ablex.

Smith, J., & Deemer, D. (2000). The problem of criteria in the age of relativism. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 877-922). Thousand Oaks, CA: Sage.

信度系数(Reliability Coefficient)

信度系数是应用一种测量方案(如一项测验、一个量表、一次专家评定等)对个体总体予以测量,测量的真实方差相对于总的观测分方差的比例的一个量度。

对于某个给定属性的任一测量操作都会给出此一属性的一个观测分,同时,在原则上,这一观测分可被分解为一个真实分【True Score】部分加一个测量误差部分。真实分被视为每个个体的一个未知常数,而误差部分被模型化为一个随机变量,同时假定它同真实分和其他误差部分都不相关。基于这些特性,代数上可以证明,对于一个给定总体,观测分同真实分的相关系数的平方等于真实分方差与观测分方差之比,亦即信度系数(参见信度【Reliability】)。因此,一个完全可信的信度评估意味着观测分的方差完全可由真实分方差抵消,并且信度系数为 1。相反,若是一个完全不可信的信度评估,观测分方差与被测属性上的实际差异毫无关系,同时信度系数为 0。虽然不存在信度系数的绝对标准,就群体比较和个体判定而言,0.7 和 0.9 被视为各自最低可接受的信度值。

到目前为止,根据观测分和真实分的变异来界定信度系数。然而,在任何测量操作后我们总可以获得观测分,但根据定义,真实分不可能直接获得。尽管如此,通过检验各次重复测量所得观测分的一致度,就可以间接地得到信度系数。检验观测分的一致度最常用的方法包括:(a)再测信度【Test-Retest

Reliability】;(b)复本信度(alternate-form or equivalent-form reliability);(c)折半信度【Split-Half Reliability】,以及(d)Cronbach α【Cronbach's Alpha】系数。

方法(a)和方法(b)处理这样的问题,将一项测验或一项测验在统计上等效的复本给同一组对象做两次,两次测验所得观测分是否是同一组观测分。而方法(c)和方法(d)解决的问题是,在某一次测验中,用于测量某个单一理论概念的两个统计上等价的经验指标组是否具有相关关系。

某些研究设计涉及由专家根据事先确定的量表或编码方案对行为进行评定。评分者信度【Interrater Reliability】和评分者内(intrarater)信度系数代表了用于评估这些测量方案信度的另外类型的信度系数。

一旦运用以上任何一种方法估出了信度系数,就可以使用它计算测量的标准误,而测量的标准误又可用以从一个人的观测分估出(点估计或置信区间【Confidence Interval】估计)他的真实分。

影响信度系数的因素包括测验的长度(即测验中题项越多,信度系数越高)、受测者的异质性(即随着被测属性上个体差异的增加,信度增加),以及测验的难度(即一个适中难度的测验可增加识别个体差异的可能性)。

——Marco Lauriola

(马妍译　赵锋校)

参考文献

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*.New York: McGraw-Hill.

重复测量(Repeated Measures)

术语重复测量指这样一种数据收集的设计方式,即在每一单元上,收集到的数据超过一个。如儿科医生每天对儿童体重予以测量,这个单元就是一个儿童;如果几年测一次通货膨胀,单元即一个国家;或者是一对夫妇,如对丈夫和妻子收入的测量。重复测量经常涉及历时的数据收集。但当前的问题在于,像许多单元的每一个观测值之间存在的那种独立性,在某一单元的多个观测值之间是不存在的。

重复测量可以仅包含两次数据收集。即配对数据设计(paired data design),而由此可以进行配对数据的 t 检验【t-Test】。丈夫和妻子的收入数据就是这种类型,就像有前测和后测设计的实验数据。针对这类数据,很容易想到的是就两个性别间的差异做一个简单的 t 检验,但这样做没有考虑数据的配对性质。两组均值差的 t 检验和配对 t 检验的分子是一致的,都等于两组均值的差值,两种检验的差别在于分母。假设夫妻中一方的收入高,另一方的收入低。这种变异会包含在标准差【Standard Deviation】之中,从而增大两组均值检验中的分母,亦即,得出一个较小的 t 值和一个更大的 p 值【p Value】。在配对检验中,重要的是丈夫和妻子间的收入差,而无须考虑原始值是高还是低。由于去除了此类变异来源,分母变得更小,t 值变得更大,相应的 p 值也变得更小和更显著了。唯一的调整在于自由度【Degrees of Freedom】。两组均值差检验有 n_1+n_2-2 个自由度,而配对检验只有 $n-1$ 个自由度。

配对 t 检验等价于双因素方差分析【Analysis of Variance】,其中每格单元只有

一个观测值,同时用于计算 F 统计量的分母不是残差均值平方,而是交互项均值平方。一个因子是性别,有两个值;另一个因子是夫妻,有 n 个值。在排除了夫妻的效应后,我们得到性别的效应。在数字上,性别的方差分析中自由度为 1 和 $n-1$ 的 F 值等于配对 t 检验的平方 t^2。

如果某个因子的类别超过两类,配对 t 检验可一般化为一个双因素方差分析,其中每格只有一个观测值,并用交互项均值平方作分母计算 F 统计量。在这种情况下,没有相应的 t 检验。若仅就一个单元在已知的时点上做多次观测,重复测量分析即变为时间序列【Time Series】分析。

——Gudmund R. Iversen

(马妍译　赵锋校)

参考文献

Cobb, G. W. (1998). *Introduction to design and analysis of experiments*. New York, Springer-Verlag.

Girden, E. R. (1992). *ANOVA: Repeated measures* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-084). Newbury Park, CA: Sage.

Hays, W. L. (1988). *Statistics*. New York: Holt, Rinehart and Winston.

凯利方格技术(Repertory Grid Technique)

凯利方格技术源于凯利的人格构念心理学(Fransella & Bannister, 1977),是一种用于表现和探索个人对事物的理解的方法。就人们先前思考很少,或很难表达的话题,请他们清楚地表明他们关于话题的知识时,这个技术特别有效。表 1 给出了一个简化的方格,它由四个特征组成:

①一个焦点:在表 1 中,焦点是在一个制造业组织中团队成员的技能。

②元素:它们是思考的对象。元素通常是人,但也可能是地点、观点或无生命的物。

③构念:它们是人们用来描述和区别不同元素的性质。构念是两极的并且具有正极和负极。

④一个连接机制:是指元素如何与构念相关。它可能是一个分类的、二分的或连续的尺度。

表 1　一个简化的凯利方格展示了一个制造业管理者对其团队成员技能的看法

	元素						
构念(正极)	熟练操作者	熟练操作者	一般熟练操作者	一般熟练操作者	非熟练操作者	非熟练操作者	构念(负极)
	Leslie	Ann	Frank	Denise	Julie	David	
节约时间	√	√	√	×	√	√	不节约时间
灵活的	√	√	√	√	×	×	不愿使用技能
质量意识	×	×	√	√	×	×	缺乏意识

续表

构念（正极）	元素						构念（负极）
	熟练操作者 Leslie	熟练操作者 Ann	一般熟练操作者 Frank	一般熟练操作者 Denise	非熟练操作者 Julie	非熟练操作者 David	
善于与管理者沟通	√	√	×	√	×	×	沟通不畅
提出观点	√	√	√	×	×	×	不愿改进绩效
有竞争性	√	×	×	×	×	×	无竞争性
自行解决问题	√	√	√	√	×	×	总是寻求帮助

资料来源：Easterby-Smith, Thorpe, & Holman(1996).

对研究者而言，首先，完成一项方格的通常方式是给出方格的焦点。其次，参与者被要求选出元素的事例，比如一个你认识的技术熟练的团队成员。构念则由"三元比较"产生，即选出三个元素，并说明同第3个相比，其中两个相似的地方。每个构念的正向和负向都是设定好的。三元比较重复进行，直至没有新的构念生成。参与者个人则通过对元素和构念间的关系的评分方式，将两者联系在一起。格子完成过程的变化包括通过简单地比较和对照两个元素（二元比较）来产生构念，还可以给出不同的元素和（或）构念。然而，有些人认为提供元素或构念有违格子的目标（即探寻"个人"的理解）。

既可通过考察构念的意义，还可通过考查构念间的关系（如构念是否可以聚集成一个更高维度［a super-ordinate factor］的因子）、元素间的关系，以及构念与元素间的关系（如哪些构念与哪些元素关联？）来分析格子。定性和定量的分析方法既可以分别使用，也可以结合使用。方格的复杂性意味着定量技术是特别有用的（Leach, Freshwater, Aldridge, & Sunderland, 2001）。然而，有学者提出定量分析有损或者违背潜在于凯利方格技术中的解释论假定，而其他人认为定量分析会忽略访谈本身所产生的丰富数据。因此，有研究者推荐对构念的意义和访谈使用定性分析。

最后，使用方格法也存在一些局限。对于所有参与者而言，它是一个耗时的过程。访谈的过程可能带来个人和敏感问题，同时，访谈者需要有技巧地化解这些问题。除非变迁本身就是一个具体的焦点，否则，凯利方格访谈并不适于阐明变化过程（如事物怎样改变和为什么改变）。

——David John Holman

（马妍译　赵锋校）

参考文献

Easterby-Smith, M., Thorpe, R., & Holman, D. (1996). Using repertory grids in management. *Journal of European Industrial Training*, *20*, 1-30.

Fransella, F., & Bannister, D. (1977). *A manual for repertory grid technique*. London: Academic Press.

Leach, C., Freshwater, K., Aldridge, J., & Sunderland, J.(2001). Analysis of repertory grids in clinical practice. *British Journal of Clinical Psychology*, *40*, 225-248.

重复(Replication)

一项社会科学实验【Experiment】的目的在于估计:(a)就一个给定总体【Population】的成员而言,某些处理对特定结果变量的效应;(b)在某些变量上,两个或更多总体之间的差异;(c)在某些特定的总体中,两个或多个变量之间的关系或相关【Correlation】。如果测量是无效的,如果样本不能代表特定的总体,如果有外来的未经确认的因素影响了因变量,那么所有这类实验都可能得出错误的结论。因此,在把任何这类实验的结论确定为一个经验事实之前,此类实验应当被重复或者复制(replicated)。

最接近字面的重复(literal replication)的一种重复,可能是由原来的作者从同一总体中抽出多个具有相同样本量的额外样本,并使用同第一次所用相同的技术、测量工具和分析方法。原初的显著性检验【Significance Testing】结果(如在抽样总体中,测量 A 和测量 B 是正相关的概率为 $p \leqslant 0.01$)即预测了(只是)这一真正字面重复的结果。只要字面重复可能达到目标,那么它实际的不可能实现就不会是很大的损失,不仅当原发现有效的时候如此,即便原发现是无效的而作者又持续地犯同样的关键错误时也如此。

当一位新的研究者尽可能原本地依照第一个作者在其方法部分提供的"实验方案",重复第一项研究时,操作重复(operational replication)即发生。一次成功的操作重复给予了某些保证,即第一位作者的发现不是某些未知变量和未控制变量的结果。

操作重复的局限性在于它的成功并不必然支持这样的宣称,例如,"在总体中,概念 X 和 Y 呈正相关",而只能说:"使用测量 A 和测量 B 得到的分数在根据某一特定方式从总体 Z 中抽得的样本中呈正相关。"这些得分可能不是对概念的好的测量,同时(或者),抽样方法可能没有产生能够代表总体 Z 的样本。此外,心理学文献的作者和编辑都学会在研究报告的方法部分,仅限于包括那些第一作者认为能够获得研究发现的必要且充分的程序细节。

最冒险也最有价值的重复形式是建构重复(constructive replication)。在建构重复中,除了特定的经验事实或宣称已经证实的事实以外,第二位作者将忽视第一份报告中的所有内容。重复者将选择他或她认为适合于研究问题的抽样、测量和数据分析方法。

一项宣称已经证实了一项理论预测的实验可以以两种方式被有效地复制:预测的效度可以被建构地重复,或者理论的递增效度可以通过实验方式检验从同一理论得出的新的且独立的预测内容被建构地重复。比如,吕肯(Lykken,1957, 1995)根据相关理论,即原发性精神病开始生活于遗传恐惧的钟形曲线【Bell-Shaped Curve】的底端处,检验了三项预测内容。有几位研究者建构地重复了原来的研究发现,但是,黑尔(Hare,1966)、帕特里克(Patrick,1994)以及其他人以实验方式证实了从同一理论得出的新的预测内容。

——David T. Lykken
(马妍译　赵锋校)

参考文献

Hare, R. D. (1966). Temporal gradient of fear arousal in psychopaths. *Journal of Abnormal and Social Psychology*, *70*, 442-445.

Lykken, D. T. (1957). A study of anxiety in the sociopathic personality. *Journal of Abnormal and Social Psychology*, *55*, 6-10.

Lykken, D. T. (1995). *The antisocial personalities.* Mahwah, NJ: Lawrence Erlbaum.

Patrick, C. J. (1994). Emotion and psychopathy: Startling new insights. *Psychophysiology*, *31*, 415-428.

定性研究的重复/可重复性(Replication/Replicability in Qualitative Research)

与信度类似,这个概念是指一个研究操作在多大程度上是一致的,可反复的。如果研究者间的一致是可达成的,就可以赋予研究发现的真实性以更大的信心。重复【Replication】指一项再研究(re-study)能再现初始研究的发现的程度。可重复性指在提供了初始研究的研究过程的充足信息的前提下,再次研究可实行的程度。

这些理念依赖于一些不为所有定性研究者共享的哲学假定。例如,认同唯心主义的研究者认为语言建构了现实,通常把他们自己的文本(研究报告等)看成他们对于研究客体的建构,而不是用传统的实在论的风格来报告真理。如果采用了这一相对主义的、社会建构论的观点,追寻重复就是没有意义的,因为每一个说明都仅仅是一个较多的"部分的真实"(Clifford & Marcus, 1986)。

即便认可实在论的假设,定性研究的重复也是困难的。消除研究者的个人特质是自然科学的世界观和定量的社会研究的某些观照的基础,而这对定性研究而言几乎是不可能的。在田野研究中,所出现的大量事物依赖于研究者的个人品质、分析的洞察力和研究者的社会技巧,而这些不可能为任何一个二次研究者所共有。定性研究的很大优势在于对特定案例的深入考察,在这些案例(有时是独特的)背景的丰富性中有所发现。案例既不会历时而不变,也不会同那些在其他层次上被断定为相同的案例有太多共同点。若干时间之后,着眼于"相同"案例的,或试图研究"相似"案例的观察者,可能实际上正在研究不同的事物,而且他们用不同的工具(即他们自身)做研究。在定性社会研究中,尝试重复一些比较著名的研究(米德和弗里曼在萨摩亚;刘易斯和雷德菲尔德在墨西哥;怀特和博伦在波士顿,参见Seale, 1999)就遇到了这些问题。现在回顾起来,这些尝试(和他们在各自领域中所引起的尴尬)又显得太过天真。

然而,可重复性是一个更现实的目标,也许,考虑到操作上的困难,它应当被重命名为"方法论会计学"(methodological accounting)。研究程序的档案文件,以及对研究项目的前提假设和个人偏好的影响的反思,都能够帮助读者评估研究者的论点所获支撑的适当性。林肯和古巴(Lincoln & Guba, 1985)概述了审核【Auditing】一个研究项目的框架,其中,为了评估项目的方法上的适当性(methodological adequacy),相应的档案文件和反思都由一队研究同行予以审查,直至研究项目的结束。

——Clive Seale

(赵锋译校)

参考文献

Clifford, J., & Marcus, G. E. (Eds.). (1986). *Writing culture: The poetics and politics of ethnography*. Berkeley: University of California Press.

Lincoln, Y. S., & Guba, E. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

Seale, C. F. (1999). *The quality of qualitative research*. London: Sage.

呈现危机(Representation, Crisis of)

短语呈现危机是指社会研究者不能直接地呈现或描绘他们所研究的人的体验【Lived Experiences】。该说法有两层含义：第一,这个词是说那些在研究者心中长期处于支配地位的,关于社会研究的性质的核心观念。关于客观性【Objectivity】、效度【Validity】、真理即符合等经典准则不再适用于我们对质询的理解。第二,为了回应这些观念的消亡,进一步的理解是,人们的体验不是由研究者所发现,并如其所是地被描述;毋宁,这些体验,如其在研究者写就的社会文本中所呈现的样子,是经由研究者在他们对他人阐释的解释中被构建出的。这一转向的含义,即从发现或找出到构建或做成,仍不断地被引申。

即使这个短语不是由马库斯和费希尔在 1986 年“杜撰”的,但确实是由于他们才引起了许多社会研究者的关注。无论如何,在此之前,这一观念即由研究者直接或客观地呈现人们的体验是一项不可能的事业,历经 20 世纪或更早的年代,经历着持续的认知上的发酵。例如,可以在 19 世纪末和 20 世纪初,各类德国学者的反实证主义著作中,在 20 世纪中叶由许多哲学家给出的领悟中——即不可能有“理论—无涉”的观察,以及在 20 世纪 70 年代和 80 年代早期的认识中——人类学的著作是阐释之阐释,社会科学和人文学科的边界已经模糊了,都可以发现这一考量的踪影(Smith, 1989)。

今天,呈现危机作为一个“危机”有着更多的历史趣味,而不再是当代的兴趣。尽管研究者可能会说,表现他者或为他者讲话仍然是个问题,但是危机感已然减弱。原因在于这样一些观念,客观性是不可能的,符合并非是一个恰当的真理的理论,不存在“理论—无涉”的知识,诸如此类,都已经获得了社会研究者的广泛认可。于是,当前的任务是要发挥和探究此类领悟的内涵。在这个意义上,诺曼·邓津和伊冯娜·林肯(Denzin & Lincoln, 2000)指出,标示出第四期定性质询的呈现危机已经过渡到第五期(后现代)、第六期(后实验的质询),以及正在发展的第七期——将道德言谈(moral discourse)作为质询的焦点。

——John K. Smith

(王玥译　赵锋校)

参考文献

Denzin, N., & Lincoln, Y. (2000). Introduction: The discipline and practice of qualitative research. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 1-28). Thousand Oaks, CA: Sage.

Marcus, G., & Fischer, M. (1986). *Anthropology as cultural critique*. Chicago: University of Chicago Press.

Smith, J. K. (1989). *The nature of social and educational inquiry: Empiricism versus interpretation*. Norwood, NJ: Ablex.

代表性样本(Representative Sample)

抽样设计的目的在于,从某个总体【Population】中选出其要素或成员的子集,以使得来自样本【Sample】的统计量【Statistics】代表总体所呈现的值。为了得到一个有代表性的样本,必须使用一种选取的概率【Probability】方法,此即,总体的每一个要素必须有一个已知的、非零的选中概率。如果这种方法使用得当,那么研究者可以确信得自样本的统计量在确定的范围内且以确定水平的信心对应于总体。

相关的总体是在具体的研究假设【Hypothesis】基础上理论界定的。总体由包括全部人口的一组要素构成,例如,"住在美国的成年人"。抽取一个概率样本的最简单的方式是有一个包含总体所有成员的清单,一个所谓的抽样框【Sampling Frame】,再把一个概率设计用于抽样框,从而生成一个代表性样本。

对于许多抽样问题来说,存在一个所有总体成员(或至少有非常接近它)的清单。为了估计上一年发生事故的司机的比例,研究者要访问司机,问他们在过去的 12 个月内是否出过事故。例如,为了生成一个司机的样本,研究者可以首先列出所有有驾照的人的清单。这个清单实际上可能包括了截至最近某日的所有有驾照的人,因此这个清单必须要更新,以包括所有那一天之后得到驾照的人。

另一个例子,研究者可能有兴趣去估计有多少美国成年人持有信用卡。因为不存在一个所有美国成年人的清单,所以这个清单必须在研究过程中被开发出来或者找到一个近似的。对那些开展电话抽样调查【Telephone Surveys】的人而言,这是一个典型的问题。在这种情况下,通常使用多阶抽样【Multistage Sampling】过程,即从电话号码的样本开始。电话号码常由计算机生成,从而解决不在电话簿上的号码和新分配的号码问题。这些号码被拨打,同时当一位住户被接通时,所有住在那里的成年人都进入清单。通过这一多阶段的过程,可以生成一个住在这一电话家庭户(telephone households)的成年人的整群抽样【Cluster Sampling】;最后一步,随机选择一个人作为回答者。这个人会被问及他或她有没有信用卡。

对于研究者来说,关键问题是样本估计值(上一年发生交通事故的司机的比例或拥有信用卡的成年人的比例)能在多大程度上代表总体的值。这是样本的两个属性的函数:它的置信区间【Confidence Interval】和置信水平。使用概率方法的一个主要的好处就是它允许研究者利用正态分布【Normal Distribution】,而该分布可以通过以某种既定方式从同一总体重复抽取样本而得到。这一抽样分布【Sampling Distribution】的特性在于包括了以下信息,分布的标准差【Standard Deviation】和样本量之间的关系,以及作为结果的样本估计值的分布。

基于上述特征,如果从一个对 1 000 名司机的简单随机样本的访问中估出,其中 14%的司机在上一年出过事故,那么将有 95%的信心说,包含所有司机的整个总体的"真值"落在 11%~17%(14%的估计值+/-3 个百分点的边缘误差)。将概率设计用于一个好的抽样框才能保证得到一个具有已知参数(即意味着它的相对精度)的代表样本。

——Michael W. Traugott

(王玥译　赵锋校)

参考文献

Kish, L. (1965). *Survey sampling*. New York: Wiley.

Levy, P. S. (1999). *Sampling of populations: Methods and applications*. New York: Wiley.

研究设计(Research Design)

一项研究项目中数据的逻辑结构构成了该项目的研究设计。它的目的是要产生能从中得出相对明确的结论的经验数据。

一项研究设计的目的在于检验和排除结果的替代解释。例如,要想检验如下命题,即婚姻会伤害女性的精神健康,而有利于男性的,就需要认真留意研究设计。一项设计可能仅仅比较已婚女性的和已婚男性的抑郁发生率。如果已婚女性的抑郁发生率高于已婚男性的,这一发现虽然同命题一致,但不会提供新的信息,以排除此发现的替代解释。如果未结婚的女性比未结婚的男性更容易抑郁,那么已婚女性的抑郁高发生率(相对于已婚男性)就很难归因于婚姻。相似地,如果婚后女性的抑郁率低于婚前的,那么已婚女性的更高抑郁水平(相对于已婚男性)就不能归因于婚姻。

一个好的研究设计能收集到可以检验每一种替代解释的数据。在经验上排除掉的替代解释越多,最后的结论也就越有根据。

大部分社会研究设计依赖于群间比较(comparing groups)的逻辑(Stouffer, 1950)。这些比较可以增强对命题的检验、对替代解释的检验,给予由数据所得结论和解释的可信性以直接的支持。例如,上述例子提到了三种不同的比较:

①已婚男性与已婚女性。

②未婚男性与未婚女性。

③婚前女性与婚后女性。

仅仅基于第一对比较的设计只能产生薄弱且含糊的结论,容许有许多替代的解释。如果设计包含有基于其他对比群体的数据,就可得出不甚含糊的结论。

不仅是所要比较的内容,而且被比较群体的性质、群组成员的招募方式,都会改变研究设计。

在社会研究中,四种主要的研究设计类型如下(de Vaus, 2001):

①实验的。

②纵贯的。

③横截面的。

④个案研究。

尽管有些作者使用研究设计一词来表述研究项目的步骤,或者是收集资料的方法(如访谈、观察等),但这是误导(Yin, 1989)。研究设计不只是收集信息的一个研究计划或一种信息——设计是一项逻辑操作,而不是一种后勤活动。这种活动类似于设计一栋建筑的建筑师的任务,而不像建立一个项目管理规划表的项目经理。

所有的研究,不论数据收集的方法,也不论它是着重于定量,还是着重于定性,都要有一个研究设计。研究设计可能是隐含的或明确的。如果它是隐含的,这些设计就经常缺乏将有用的比较群体纳入研究的意识,这样从数据中得到的结论就缺乏有力支持。

——David A. de Vaus

(王玥译　赵锋校)

参考文献

de Vaus, D. A. (2001). *Research design in social research*. London: Sage.

Stouffer, S. A. (1950). Some observations on study design. *American Journal of Sociology*, *55*, 355-361.

Yin, R. K. (1989). *Case study research: Design and methods*. Newbury Park, CA: Sage.

研究假设(Research Hypothesis)

做经验研究是为了认识我们周围的世界。它源于一般的关于行为的直觉和理论——社会的、生物的、生理的,诸如此类。大部分研究关注变量之间可能的关系,而直觉和理论要转变成关于变量的具体陈述。此类说明即陈述研究者认为变量是如何相联系的研究假设。因为不知道这些陈述是真还是假,所以可将它们当作对世界的提问。然而研究者可以希望通过收集和分析经验数据获得答案。

"共和党人比民主党人赚的钱多吗?"可能就是一个属于更大的政治行为理论的问题,同时,可能的研究假设是,"共和党人比民主党人赚钱多。"可以进行一次抽样【Sample】调查,从而获得人们的投票行为和年收入的数据。

一个研究假设需要转成一个统计虚无假设【Null Pypothesis】。统计推论系统不能直接使用研究假设。因此,研究者要确定一个或多个统计参数,并给出参数【Parameters】的可能值,从而替代研究假设。这就产生了一个统计虚无假设,可以使用统计推论的方法分析它,即使用经典的统计假设检验【Hypothesis Testing】。贝叶斯【Bayesian】方法可以避免一些与频次论假设检验相关的难题。

把研究假设转成一个更具体的统计虚无假设可能存在困难。用我们的例子,一个可能的虚无假设是共和党人的人均收入与民主党人的人均收入相等,即 $H_0: \mu_r = \mu_d$。如此一来,原来的问题就变为关于均值的问题,而这可能是也可能不是研究者所想回答的。

同样,统计虚无假设说,两个群体的均值是相等的,而研究假设说,一个群体在某种程度上高于另一个群体。虽然研究假设指出,一个群体比另一个群体赚得更多,但是虚无假设的焦点却相等。其原因是:大部分研究的目标是要证明无差异或无关系的虚无假设是错误的。如果可以排除两个均值相等的可能,我们就已经表明它们一定不同。这就是以一个双重否定的方式来证明研究假设为真。

即使数据表明一个研究假设是真的,数据仅仅涉及变量的统计关联或差异,并不涉及变量间的因果【Causal】模式。要建立因果关系(causality)是极其困难的,而且绝大多数情况下,它只限于实验数据。观测数据本身不易确定变量间的因果关系。基于观测数据建立因果关系的稀少例子之一是每包香烟上的告诫:吸烟有害健康。

——Gudmund R. Iversen

(王玥译 赵锋校)

参考文献

Bickman, L., & Rog, D. J. (1998). *Handbook of applied social research methods*. Thousand Oaks, CA: Sage.

Crano, W. D., & Brewer, M. B. (2002). *Principles and methods of social research* (2nd ed.). Mahway, NJ: Lawrence Erlbaum.
Henkel, R. E. (1976). *Tests of significance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-004). Beverly Hills, CA: Sage.

研究管理(Research Management)

成功管理一个研究项目或一个研究机构包含很多方面。本质上,一名研究管理者的首要目标是利用可用的资源,最高程度地提高一个研究或一个研究机构所收集到资料的质量(例如信度【Reliability】和效度【Validity】)。这个过程既不简单也不容易,占据研究管理位置的人如果让富有经验的研究者根据他们自己的权利从事研究,将受益匪浅。归根结底,研究管理者必须能够就许多研究设计和研究操作上的妥协——在任何研究的计划和执行中都会遇到的——帮助作出明智的决策。此即,就任何研究而言,仅有有限的资源可用,因此,资源必须得到明智的配置(从成本效益的立场)。

除了对于即将开展的研究方法论和统计方面有经验和专长,一个成功的管理者或管理团队很可能还需要募集资金、财务、人事管理、沟通和公共关系,以及研究伦理的知识和技能。没有充分的财务和人力资源,一个研究项目就不可能收集到能够满足研究目标的相应质量的数据。因此,需要募集资金的技能,以吸引可用于收集有效数据【Data】和可用于分析、解释和传播研究发现的必要资源。要维持正常的研究工作,管理者还要负责研究成员的招募、雇用、激励、评估和挽留。沟通技能也是必需的,这样研究产生的新知识才能得到传播(很多时候,研究发现的传播有助于吸引未来的资金募集)。管理者还必须慎重对待研究中的伦理问题,尤其是那些涉及人、动物的健康和/或环境问题,以及那些需经机构审查委员会核准的问题。

——Paul J. Larrakas
(王玥译　赵锋校)

研究问题(Research Question)

如果一个研究议题具有可考察的形式,其内容和范围也得到界定,就成为一个研究问题。正是通过使用这些问题,才能作出关于研究的关注点和方向的选择,才能清楚地界定研究的边界,才能使得研究具有可操作性,也才能期望有成功的结果。一个或多个研究问题提供了一个研究计划得以成立的基础。它们的选择和措辞决定了要研究什么、研究到什么程度,以及怎样研究。

研究问题的主要类型有三种:是何(what)、为何(why)和如何(how)。"是何"问题需要一个描述性的答案;这些问题引导研究者去发现和描述某个社会现象的特征和规律(如个体的类别、社会群体的各种规模以及社会过程)。"为何"问题要求回答特征或规律存在的原因或理由。它们关注的是理解或解释事件之间的关系,或社会行动和社会过程内部的关系。"如何"问题关注的是引起改变,以及实际的后果和介入(Blaikie, 2000, pp. 60-62)。

这三种类型的研究问题形成一个序列："是何"问题通常在"为何"问题之前，"为何"问题常在"如何"问题之前。在我们能作出解释前，要知道发生了什么；此外，在我们能有信心提出一种介入手段去改变它之前，要知道它为何以此方式发生作用。然而，大部分研究项目仅包括其中的一类或两类研究问题。最常见的是，"是何"和"为何"问题，甚至有些研究只处理"是何"问题（Blaikie，2000，pp. 61-62）。

有些作者举出了超过这三种类型的研究问题，如"何人""是何""何处""何多""何甚""如何""为何"（Yin，1989）。然而，除了"为何"和"如何"问题是真正的不同类型，其他所有类型都可以换成"是何"问题："何人"在"何地"、在"何时"、以"何其多的数或量"，以及"何种方式"。

依据研究项目在起始阶段产生精确的研究问题的可能性，研究项目会有所不同。在研究前，首先开展一些探索性的研究可能是必要的。另外，在实施研究过程中，可能要根据发现的内容，不时地重新考虑研究问题。

系统地陈述研究问题是所有研究设计【Research Design】中最关键的，也许是最困难的部分。然而，很少有研究方法的教科书花精力去讨论它们，有些教科书甚至完全遗漏了研究过程的这一核心部分；真正讨论它们的人则偏向于定性研究（Flick，1998；Mason，2002）。

传统的见解认为研究应该由一个或多个假设来引导。这种立场的支持者认为，研究当始于问题的系统陈述，随之以一个或多个假设的设定、概念的测量，以及对所测变量的分析以检验假设。然而这个过程只切合于使用演绎【Deductive】逻辑的定量研究。尽管在此类研究中假设有着重要的角色，但它们既没有为研究设计提供基础，对于界定研究问题的关注点和方向也没有帮助。

在某些类型的研究中，或者没有可能，或者没有必要，从假设出发。所需要的是建立一个或多个研究问题。假设的作用在于给出"为何"问题，有时也包括"如何"问题的尝试性回答。它们是对答案的最好猜测。然而，它们不适于"是何"问题。冒险去猜测事物的可能状态是没有依据的。随着研究的开展，一段时间之后，研究自会给出"是何"问题的答案，而就将要发现的内容的所有猜测对真正的发现不会有所帮助（Blaikie，2000，pp. 69-70）。

——Norman Blaikie

（王玥译　赵锋校）

参考文献

Blaikie, N. (2000). *Designing social research: The logic of anticipation*. Cambridge, UK: Polity.

Flick, U. (1998). *An introduction to qualitative research*. London: Sage.

Mason, J. (2002). *Qualitative researching* (2nd ed.). London: Sage.

Yin, R. K. (1989). *Case study research: Design and method* (rev. ed.). Newbury Park, CA: Sage.

残差(Residual)

残差用于量度任一估计量的实测值与预测值之间的距离。如果 $\hat{y}$（读为 y）是 y 的估计值，那么可以简单地用 y 的估计值减去它的真实值来计算残差（$\hat{u}$）：$\hat{u}=\hat{y}-y$。残差

可有多种使用方式。它们能被用于计算回归【Regression】线、估计拟合优度【Goodness-of-Fit Measures】,还可以作为数据分析中的一种诊断技术。

最小二乘【Ordinary Least Squares,OLS】用平方的残差进行回归估计。OLS会拟合一条对数据而言其残差平方和具有最小值的回归线。因此,OLS要生成一条具有尽可能小的残差的回归线。同时,由OLS导出的正规方程组(normal equations)会生成具有特定属性的残差。首先,残差的期望值为0,但这并不意味所有残差值为0,而是所有残差的平均数为0。其次,每个自变量【Independent Variable】与OLS残差的样本协方差【Covariance】为0。此外,OLS残差与拟合值的样本协方差也为0(Wooldridge,2003)。

残差也用于估计回归模型的拟合优度。通过判断实际观测值与估计的回归线之间的接近程度来评估拟合的优越;即实际观测值越接近,拟合得越好。可用残差计算几种不同的拟合优度指标,其中R方【R-Squared】和均方根【Root Mean Square】是两种不同的用残差计算的拟合优度。

残差也可以告诉人们数据和模型的很多情况。在数据分析过程中,会多次检验残差。可以通过检视它们同回归线的残差图来实现。如果残差是负的,可得知回归方程高估了因变量【Dependent Variable】的值;如果残差是正的,那么回归线低估了因变量的值。用这种方式检验残差还可以知道,是否存在某些用模型不能很好予以拟合的观测值类型。达摩达尔·古雅拉提(Damodar N. Gujarati,2003)提出残差图是一种有效实践。在横截面数据【Cross-Sectional Data】中,残差图可以反映出模型的误设【Misspecification】和异方差【Heteroskedasticity】问题。如果一个模型缺失了一个重要变量,或者设定了错误的函数形式,那么残差将呈现出独特的模式。在时间序列【Time Series】中,检验残差是一种可识别出异方差和自相关【Autocorrelation】的有效实践。

——Heather L. Ondercin

(马妍译　赵锋校)

参考文献

Gujarati, D. N. (2003). *Basic econometrics* (3rd ed.). New York: McGraw-Hill.

Wooldridge, J. M. (2003). *Introductory econometrics: A modern approach*. Cincinnati, OH: South-Western.

残差平方和(Residual Sum of Squares)

方差分析【Analysis of Variance, ANOVA】、线性【Linear】和多元回归【Regression】的一个关键成分是残差平方和,也称为误差项平方和或误差平方和。残差平方和是因变量观测值和预测值之间的差异,反映了因变量上的未由线性回归模型解释掉的方差。

在最小二乘回归中,将每一个残差的值取平方,再加总所有观测到的残差平方,即得到残差平方和,记作SSE。残差平方和表示未被线性模型解释掉的平方和。它描述了观测值到预测线的变差。以数学来表示,残差平方和(SSE)即 $\sum (Y_i - \hat{Y}_i)^2$,其中,Y_i = 观测值,$\hat{Y}_i$ = 预测值。

下面给出了一个数值例子,说明如何使用残差平方和。在一项包含四个人的体重

减轻的假想研究中，因变量是体重的减少量（Y），自变量是运动（X）。四个人中的每一个人在2周中减轻的体重数（以磅为单位）都反映在表1的Y列中。X列反映了观测到每个人在2周的课程中完成的运动量（以小时为单位）。$\hat{Y}$列是根据预测线得到的Y的预测值。$(Y-\hat{Y})$列表示Y的每个观测值和预测值之间的个体残差。最后，$(Y-\hat{Y})^2$列给出了残差的平方。将所有残差的平方相加即可得到残差平方和。在这个例子中，其值为677。

表1 残差平方和示例

观测值	X	Y	$\hat{Y}$	$(Y-\hat{Y})$	$(Y-\hat{Y})^2$
1	7	28	14	14	196
2	15	12	30	−18	324
3	8	5	16	−11	121
4	10	14	20	−6	36

残差平方和可用于衡量回归预测的精度。随着预测方程的改善，即预测值更接近于观测值，每个观测值的残差都趋向减少，相应地，残差平方和也变小。

残差平方和可以用来确定决定系数【Coefficient of Determination】，在两变量模型中记作r^2。该系数表示使用线性预测方程代替样本均值来预测因变量的值，从而获得误差消减的比例。决定系数的方程是(TSS−SSE)/TSS，其中，TSS＝总平方和，SSE＝残差平方和。就多元回归而言，可以使用同样的公式计算多元决定系数，记作R^2。

在确定决定系数的意义时，残差平方和有重要作用。相关的检验统计量被称为F比率【F Ratio】，其计算公式为：

$$\frac{\dfrac{R^2}{k}}{\dfrac{1-R^2}{n-(k+1)}}$$

其中，R^2为决定系数；k为模型中解释变量的数量；n为样本量。更进一步，在多元回归中，残差平方和可用来比较完整模型和简化模型。该检验统计量也是F检验，其计算公式为：

$$\frac{\dfrac{SSE_r-SSE_c}{k-g}}{\dfrac{SSE_c}{n-(k+1)}}$$

其中，SSE_r为简化模型的残差平方和；SSE_c为完整模型的残差平方和；k为完整模型中解释变量的数量；g为简化模型中解释变量的数量；n为样本量。这个检验允许使用者确定，就精确地预测因变量的值而言，哪个模型更好。

——Shannon Howle Schelin

（马妍译　赵锋校）

＊也可参见误差平方和【Sum of Squared Error】。

参考文献

Agresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Lewis-Beck, M. S. (1995). *Data analysis: An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-103). Thousand Oaks, CA: Sage.

回答人(Respondent)

回答人即被选作回答调查问题的人。在许多情况下,在抽样框【Sampling Frame】中的抽样单位、回答人,以及研究者想要报告的研究总体【Population】的成员或单位之间,存在着一一对应的关系。在这种情况下,研究者可以从符合参与研究项目条件的研究总体的个体成员清单中选出一个样本。这时,抽样过程即识别回答人的过程。例如,在调查校长的态度和行为时,通过对学校清单的抽样,可以将特定的公立学校的校长认作回答人,因为,一般而言,每所学校都有且只有一位校长。这时,抽样单位、回答人和分析单位【Unit of Analysis】即完全一致。

回答人可能与抽样单位不同的原因有三点:

1. 抽样单位是个体分析单位的群组或集合。

2. 调查实施的方式不允许回答人直接回答问题,即使回答人有能力提供答案。

3. 研究总体的部分或全部没有能力作出回答。

下面就上述三点原因及其解决方法展开陈述。

抽样单位是集合

许多调查,如那些使用随机数字拨号【Random Digit Dialing】的电话抽样调查或区域概率抽样【Probability Sampling】的个人访谈(Henry, 1990; Lavrakas, 1987),首先是找出家庭户或其他形式的个体组合,再从中选取一个具体的回答人。如果选取那些接电话或应门的人作回答人就会产生有偏差的样本,因为某类家庭成员(通常是女性)比家庭中的其他人更可能作出上述回应。因此,社会科学家设计了几种客观的程序,以选出适合的家庭成员(通常是18岁及以上,且住在家里的成员)回答调查问题。最常使用三种方法来极小化系统偏差:Kish法、Trodahl-Carter法和最近生日法(Czaja, Blair, & Sebestile, 1982; O'Rourke & Blair, 1983)。例如,全国选举研究(National Election Studies,NES)使用Kish法,通过一定的抽样程序,从每一个选中的家庭户中,选出一名具有投票资格的人作为回答人。一旦有投票资格的回答人被选中,NES的访问程序则不允许替换。

在某些情况下,可以依据某些对研究而言重要的标准,系统地选出回答人,如选择那些最了解家庭情况的成员或家庭中的一位特定成员。这种情况经常出现在回答人被问的不是与他或她自己有关的态度和行为,而是与其他个体或整个集体(如家庭)有关的事项。例如,赫斯(Hess,1985)指出:

> 对消费者财务状况而言,比较偏爱的回答人是那些关心财务数据汇集的人,也就是财务支出单位的户主;仅当访谈持续期,无法接触到户主时,才对户主的妻子进行访谈。(p. 53)

调查程序排除了某种回答人

有些调查的实施预先就将样本研究总体成员的某些人排除于应答之外。比如,用英语展开的访谈事先排除了不说英语的回答人。这种情况下,研究者可以作出选择,或者让代理人来回答问题,例如,如果可能,请一个说英语的家庭成员来为那位不说英语的研究总体成员做翻译,或者对不说英语的人不予访谈。任何一种选择都可能使回

答发生偏差,因此,研究者必须决定哪一种方式可以让研究偏差最小,或是必须考虑用其他非英语语言做调查。调查的局限不仅体现在语言的差异上,也体现在研究总体成员的听力障碍或表达障碍上。

研究总体成员不能回答

通常,研究总体成员或者某些被选作样本的成员不能回答调查问题。例如,对幼小儿童的研究经常要依靠他们的父母或老师来提供回答。某些情况下,样本的目标总体成员中的一些人,如老年人或严重疾病患者,可能不能回答问题。研究者必须决定是否允许其他人代替或帮助他们回答,或是将那些不能为自己提供答案的应答全部省去,并将之归入无回答偏差【Nonresponse Bias】。

在大多数调查中,如果存在样本清单,而且抽样单位就是回答人,就不需要额外的程序来指定回答人。然而,社会科学中最常用的两种抽样形式,区域概率抽样和随机数字拨号都需要额外的程序选取回答人,因为这两种抽样程序选出的只是家庭户,而非个体。还有些时候,调查研究者面临着一些调查总体成员,他们虽然是有被调查的资格,但是由于没有某种类型的临时居所而被排除于调查之外。近年来,英语国家中不讲英语的人的比例在增加,而调查者也已经投入了许多努力来增加那些可能提供调查应答的个体的数量。社会研究者既要决定对那些临时居民实施调查,还要考虑针对临时居民发展相应调查程序、翻译调查工具和实施调查的诸项成本和能力。

——Gary T. Henry

(王玥译 赵锋校)

参考文献

Czaja, R., Blair, J., & Sebestile, J. P. (1982). Respondent selection in a telephone survey: Comparison of three techniques. *Journal of Marketing Research*, *21*, 381-385.

Henry, G. T. (1990). *Practical sampling*. Newbury Park, CA: Sage.

Hess, I. (1985). *Sampling for social research surveys 1947-1980*. Ann Arbor, MI: Institute for Social Research.

Lavrakas, P. J. (1987). *Telephone survey methods*. Newbury Park, CA: Sage.

O' Rourke, D., & Blair, J. (1983). Improving random respondent selection in telephone surveys. *Journal of Marketing Research*, *20*, 428-432.

回答人确认(Respondent Validation)

参见成员确认和审核【Member Validation and Check】。

回答偏倚(Response Bias)

在一份心理学评定量表和自我报告中,一个回答偏倚是一个人在他或她的判断中带入特殊的系统性偏倚的一种倾向。这些偏倚引入了主观性,会歪曲判断。如果这类偏倚【Bias】同时出现在不同变量的测量中,它们将扭曲变量间被观测到的关系(参见方

法方差【Method Variance】)。应当注意,人们的倾向会改变,从而展现出不同的回答偏倚。

在涉及自我的量表中,研究最充分的回答偏倚或许是社会期许【Social Desirability】。这是在回答有关自我的题项时,不顾真实情况,而循着某种社会期许的方向作答的倾向。社会期许回答偏倚只会出现在那些具有社会敏感性的题项上,而不会出现在所有题项上。例如,人们通常更可能在心理调试的评定中而不是工作满意度的评定中看到这种偏倚(Spector, 1987)。除了回答偏倚,社会期许也被视为需要获得认可的人格变量(Crowne & Marlowe, 1964)。那些对获得他人的认可有高度需要的个体倾向于依照某种社会期许的方向回答问卷。

在涉及对别人进行判断的评定量表中,最常见的回答偏倚是荣光(halo),即就同一个体的所有维度予以同样评定的倾向,而不顾真实存在的变异。因此,呈现这种偏倚的个人会就某个人在所有维度上给出高的评定,而给另一个人的所有维度低的评定,即便每个人在诸测量维度上可能是有变化的。这可能是由于,评估者先对每个人做了一个总判断,再假定这人在所有测量维度上都符合总判断。这类回答偏倚与反应定势【Response Sets】(比如,宽容和严格)不同,因为后者不依于评定的题项或评定的对象。例如,教师在给学生评分时,一种反应定势可能是倾向于给所有的学生相同的评分(比如,宽容地给所有学生高分)。而回答偏倚可能是倾向于在对每个人的评分中加入主观性,给喜欢的学生偏高的分数,给不喜欢的学生偏低的分数。

评定偏倚可以根据参与者的回答模式进行推断,但不总是可能精确地将偏倚和准确的判断区别开。例如,就荣光偏倚而言,人们已经认识到,在许多情况下,不同的测量维度间可能存在真实的关系。在员工工作绩效领域,研究者已从包含了评级偏倚的虚假荣光(illusory halo)中区分出了真实荣光(true halo),或者说工作表现的诸测量维度间的真实关系(Latham, Skarlicki, Irvine, & Siegel, 1993)。

——Paul E. Spector

(马妍译　赵锋校)

参考文献

Rohrer, L. G. (1965). The great response-style myth. *Psychological Bulletin*, *63*, 129-156.

Spector, P. E. (1987). Method variance as an artifact in selfreported affect and perceptions at work: Myth or significant problem? *Journal of Applied Psychology*, *72*, 438-443.

回答效应(Response Effects)

回答效应是一个一般的术语,涵盖了一系列对抽样调查【Survey】问题的响应,而这些响应独立于调查所问问题的目的。因此,像某个问题的背景效应(即调查中相关问题的其他类型是什么),顺序效应【Order Effects】(即回答的问题位于相关问题之前还是之后),记忆效应(即从问题到问题,回答者有多好的记忆),措辞效应(即变换问题的措辞能否引发出不同的回答),社会期待【Social Desirability】效应(即回答者给出

的是他们认为"正确"的答案吗),以及反应定势【Response Set】(即不管问题是什么,都喜欢一个特定的回答选项),都是回答效应的例子。从一定意义上讲,这类回答效应都是误差的来源。

——Michael S. Lewis-Beck

(王玥译 赵锋校)

反应定势(Response Set)

在一项心理学自我报告或评定量表中,反应定势是不依于所问的问题或需要判断的刺激(人),而表现出某种特定反应模式的倾向。例如,就一组多项选择的测验题,反应定势可能是喜欢选取某个特定选项,如选"3"或者"C"。

存在着几种可辨识的反应定势。在自我报告的量表中,默许(acquiescence)是指无论题项内容如何,都说"是"(是或否题)或表示同意(同意或不同意题)的倾向。极端情况下,展现出此种定势的人可以对所有赞同和不赞同题项都表示赞同,即使这意味着他或她的答案存在矛盾。也可能存在非默许(nonacquiescence)定势,即无论题项内容如何,该人都表示不同意。

在人们被要求对他人行为的几个维度作出评价的量表中,发现存在 3 种反应定势,比如评估员工的年度表现,或者评估申请研究生学习的学生能力。宽大(leniency)是在测量的目标维度上,给所有个体以高评的倾向。例如,可以发现,一个评定人倾向于只用一个五级评定量表中最高的两个选项,如教师在打分时仅使用"A"和"B"两个等级。严格(severity)正好相反,评价者仅给予低评,如教师只给学生"D"等或"F"等成绩。中等倾向(central tendency)是只使用量表的中间选项的倾向,如老师倾向于给大多数人"C"等成绩,而很少人得到"A"等或者"F"等。

虽然在理论上反应定势是不依于所问题项或被评判的事物的真实水平的反应模式,但是这种定势似乎并不是真正地不依赖于内容。例如,罗勒(Rohrer,1965)能够证明,默许反应定势并非不依赖于内容,而是出现在同一测量工具的诸题项上,不会同时出现在不同的测量工具上。换句话说,一个就某一测量工具表现出某种反应定势的个体,就其他测量工具则不会显出同一倾向。斯佩克特(Spector,1987)表示,对于赞同和不赞同的题项都表示赞同的倾向非常少见。人们更可能展现回答偏倚【Response Bias】,即他们回答的模式会受到所问题目和被判断对象的影响。

——Paul E. Spector

(马妍译 赵锋校)

参考文献

Rohrer, L. G. (1965). The great response-style myth. *Psychological Bulletin*, *63*, 129-156.

Spector, P. E. (1987). Method variance as an artifact in selfreported affect and perceptions at work: Myth or significant problem? *Journal of Applied Psychology*, *72*, 438-443.

回溯推理(Retroduction)

在哈雷(Harré, 1961)和巴斯卡尔(Bhaskar, 1979)的实在论【Realism】观点中,以及那些被称作超越的或批判实在论【Critical Realism】观点中,可以发现这一质问的逻辑。哈雷和巴斯卡尔认为,回溯推理被认为能够克服归纳和演绎逻辑的缺陷,从而给予因果解释。他们还主张,它曾是过去重大的科学解释的基础。

回溯推理伴随着这样一种观念,即从已发现的模式或规律性退回到它们的下面或后面,以发现产生它们的事物。哈雷和巴斯卡尔都拒绝实证主义【Positivism】的解释模式,即可通过设立规律或恒定的联系,而得到解释。对他们来说,建立这类规律只是研究过程的开始。之后需要的是去确定产生规律的结构或机制。这些结构和机制是事物以特定方式表现的趋势或力量。事物表现其力量的能力或它可能表现其力量的可能性,依赖于环境是否有利。

根据巴斯卡尔(Bhaskar, 1979),新的机制属于一个尚未被观察到的实在的领域。为了发现这些机制,就必须建立它们的假设模型。如果这些机制存在并以假定的方式表现,就可以用它们说明被考察的现象。这些模型构成假设的描述,而这些描述,依其希望将揭示潜在的机制。此即,只有先构建关于这些机制的想法,它们才能被认识。回溯推理等同于这一过程。

建立这些假设模型是一项创造性的活动,既需要经过训练的科学想象力,也需要运用类比和隐喻。一旦构建起了模型,研究者的任务则要去确定假定的结构或机制是否实际存在。这将涉及假设机制存在,以检验它的预测力,或者发明新的工具去观测它。假设模型的主要价值在于它给研究指明方向;不像归纳性的研究者,回溯推理(retroductive)的研究者已经知道要去找什么。

回溯推理的逻辑已经在自然科学中用于发现过去未知的因果机制,如原子和病毒。哈雷将其用于他的行为发生学【Ethogenic】方法以研究社会心理学(参见Harré & Secord, 1972);波森则结合哈雷和巴斯卡尔的观点,将其变成社会解释的一种实在论形式(Pawson & Tilley, 1997)。布莱基(Blaikie, 1993, 2000)评论了回溯推理的观念及其在社会科学中的应用。

因为回溯推理是一个社会科学中相对新的观念,它还没有引起严肃的批评。然而,由于以批判实在论为基础,所以对后者的批判也将指向回溯推理这一概念。在大多数情况下,这些批评很可能来自其他范式【Paradigm】的信徒。

——Norman Blaikie
(王玥译　赵锋校)

参考文献

Bhaskar, R. (1979). *The possibility of naturalism*. Brighton, UK: Harvester.

Blaikie, N. (1993). *Approaches to social enquiry*. Cambridge, UK: Polity.

Blaikie, N. (2000). *Designing social research: The logic of anticipation*. Cambridge, UK: Polity.

Harré, R. (1961). *Theories and things*. London: Sheed & Ward.

Harré, R., & Secord, P. F. (1972). *The explanation of social behaviour*. Oxford, UK: Basil Blackwell.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: Sage.

修辞学(Rhetoric)

修辞学一词既指言语的说服特征,又指长期确立的研究演讲术的传统。例如,政治演说,即修辞学在第一种意义下的例证,因为它们是有目的地劝导他人的交流。同样,人们可根据修辞训练来研究政治演说,其目的在于示范,演讲者如何使用论说工具以期说服听众。

对修辞学的研究可以追溯到古代。例如,亚里士多德和西塞罗分析了雄辩术的性质,并试图为有志成为公共演说家的人提供系统的、实践的指导。在中世纪的欧洲,修辞学与数学和语法并列,是学校课程的三个基本学科。到了文艺复兴时期,修辞学成为对书面文学的研究,与此同时,修辞学家付出很多努力编纂语言的各种修饰应用准则。直至19世纪末,新的语言学、心理学和社会学开始取代修辞学在语言和说服研究中的显要地位。

经典的修辞学本质上是前方法论的,因为它没有给予分析者一套现代意义上的方法论程序。这是何以修辞学传统似乎走向衰亡的原因之一。然而,思想家(如伯克[Kenneth Bunke]和佩雷尔曼[Chaim Perelman]),使经典修辞学传统在20世纪中叶得以复苏。这一新的修辞学对人文和社会科学有一个非常重要的影响——新修辞学家拓展了修辞学,使其超出传统的演讲术和文学文本的传统论题,涵盖所有言语的研究,包括学术和科学的写作。

尤其是"修辞学质询"试图证明科学和社会科学是深度的修辞事业(Nelson, Megill, & McCloskey, 1987)。甚至是那些标榜自己为"事实的(factual)"和"科学的(scientific)"学科也在他们的技术报告中使用修辞工具。例如,麦克洛斯基(McCloskey,1986)在她的重要著作中《经济学的修辞术》(*The Rhetoric of Economics*)将伯克的修辞学理论用于经济学。她认为经济学家习惯用形象的说话方式,如隐喻和借代,而且,他们征用数学公式作为说服的装饰工具。经济学家使用此类工具让读者相信他们结论的事实性和科学性。其他一些研究者考查了心理学、社会学、数学和生物学著作中的修辞惯例。

质询的修辞学显示,在人类事务中修辞术无处不在(Billig, 1996)。对学术著作中修辞术的揭露,不可避免地对那些公开宣称自己是事实的,且因此据信是非修辞的学科具有批判意味。而且,质询的修辞学证明,科学和社会科学不仅由方法论和理论构成,写作实践也是关键,尽管其实践者常常不刻意于此。迄今,就写作实践,也就是个别学科的特有修辞术,影响理论内容及方法实践的程度仍在讨论中。

——Michael Billing

(王玥译 赵锋校)

参考文献

Billig, M. (1996). *Arguing and thinking*. Cambridge, UK: Cambridge University Press.

McCloskey, D. N. (1986). *The rhetoric of economics*. Madison: University of Wisconsin Press.

Nelson, J. S., Megill, A., & McCloskey, D. N. (Eds.). (1987). *The rhetoric of the human sciences*. Madison: University of Wisconsin Press.

总体相关系数(RHO)

总体相关系数是一个总体参数,表示皮尔逊相关系数【Pearson's Correlation Coefficient】的总体相关系数。设想有一个包含 N 个个案的总体,同时,在随机变量【Random Variable】X 和随机变量 Y 上,每个个案都有一个取值。换言之,存在着一个包含有 N 对(X, Y)的总体。变量 X 和变量 Y 之间的总体相关系数(记作 ρ_{XY})定义如下:

$$\frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

或者,

$$\frac{\sum z_{X_i} z_{Y_i}}{N}$$

两者在数学上是等价的。σ_X^2 和 σ_Y^2 是变量 X 和变量 Y 的真实方差,可以通过以下公式获得:

$$\sigma_X^2 = \frac{\sum (X_i - \mu_X)^2}{N}$$

和

$$\sigma_Y^2 = \frac{\sum (Y_i - \mu_Y)^2}{N}$$

其中 μ_X 和 μ_Y 分别是变量 X 和变量 Y 的总体均值。Cov(X, Y)是总体的真实协方差,定义为:

$$\frac{\sum (X_i - \mu_X)(Y_i - \mu_Y)}{N}$$

z_{X_i} 和 z_{Y_i} 是个案 i 在变量 X 和变量 Y 上的 z 分数【z-Scores】,定义为:

$$z_{X_i} = \frac{X_i - \mu_X}{\sigma_X}$$

和

$$z_{Y_i} = \frac{Y_i - \mu_Y}{\sigma_Y}$$

总体相关系数可能在 $-1 \sim 1$ 变动。ρ_{XY} 的正值表示,变量 X 的一个增量伴随有变量 Y 的一个增量,反之亦然。相反,ρ_{XY} 的负值表示,变量 X 的一个增量伴随有变量 Y 的一个减量,反之亦然。ρ_{XY} 取值为 0 表明变量 X 和变量 Y 之间没有线性关系。要注意,ρ_{XY} 取值为 0 只是变量 X 和变量 Y 相互独立的必要条件而非充分条件。因此,ρ_{XY} 取值为 0 并不意味着变量 X 和变量 Y 一定相互独立。然而,如果两个变量是相互独立的,那么 ρ_{XY} 必然等于 0。

从总体中选取的样本的相关系数(r)通常与它的总体相关系数(ρ)不同。试想一下,随机地从一个总体中抽选了一个样本【Sample】,其中有 20 对(X, Y),并给定 $\rho_{XY} = 0$,再根据以下公式计算皮尔逊相关系数:

$$\frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sqrt{\left[\sum X_i^2 - \frac{(\sum X_i)^2}{n}\right]\left[\sum Y_i^2 - \frac{(\sum Y_i)^2}{n}\right]}}$$

运用相同的程序,可以重复 1 000 次抽选出包含 20 对(X, Y)的样本,再分别抽选包含 40 对、80 对、160 对、500 对和 1 000 对(X, Y)的样本各 1 000 次。六种不同样本

量的 1 000 次抽样的相关系数的频次分布显示在图 1 中。虽然六个分布总体上都是对称的而且彼此相似，但是随着样本量的缩小，样本相关系数的全距在扩大。正如表 1 所示，当样本量从 20 扩大到 1 000 时，样本相关系数的全距从[−0.71，0.72]下降到[−0.08，0.09]。此外，给定 $\rho_{XY}=0$，当样本量扩大时，样本相关系数的均值更趋近于接近总体相关系数。当 ρ_{XY} 等于非 0 取值时，这些分布之间的比较将会在皮尔逊相关系数中得到详细描述。

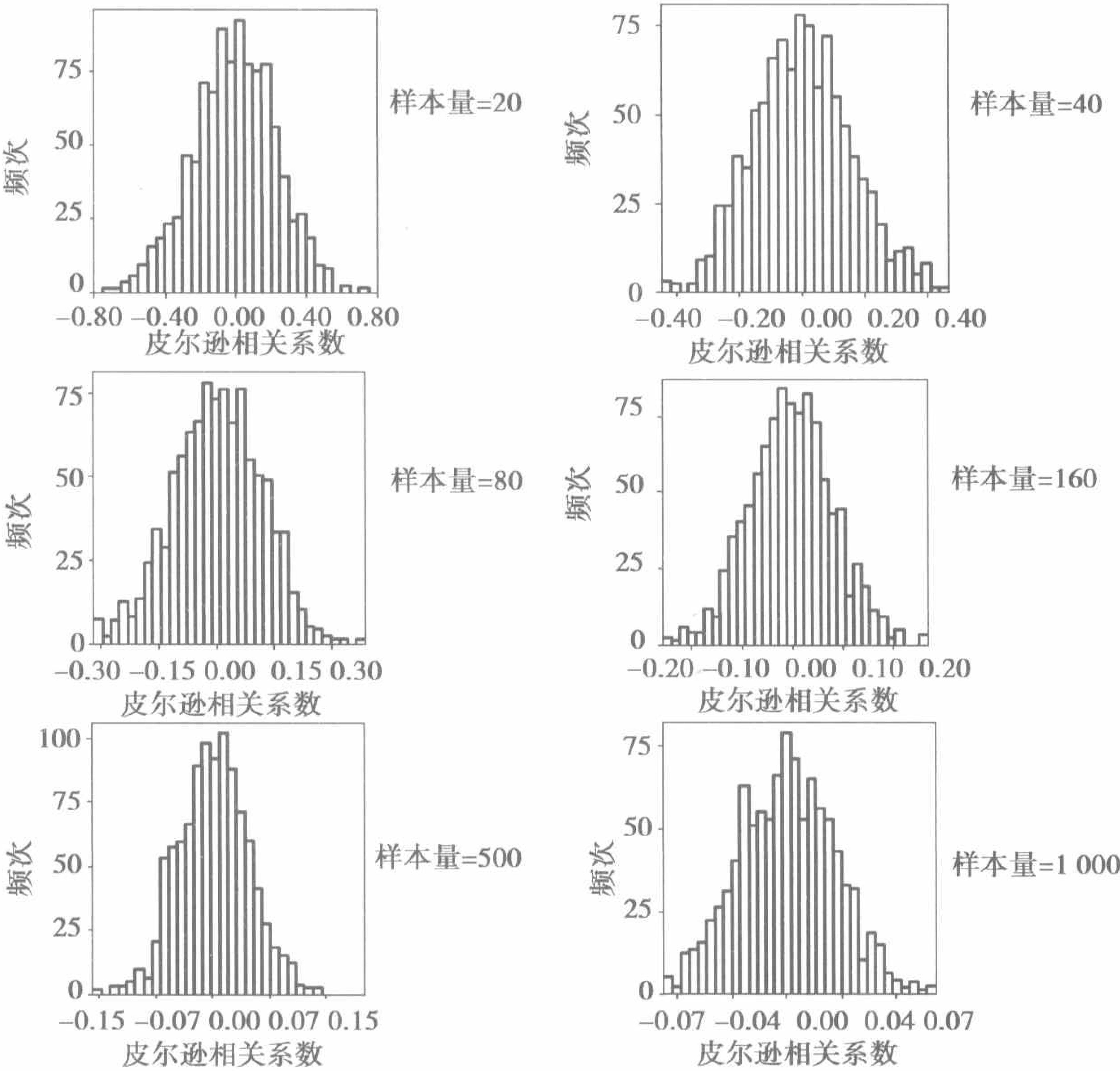


图 1　给定 $\rho_{XY}=0$，六个不同样本量的样本相关系数的频次分布

表 1　基于六种样本量的蒙特卡洛模拟的描述结果

样本量	标准差	最小值	最大值	全距	第 20 位百分位数	第 50 位百分位数	均值	第 75 位百分位数
20	0.229 620 2	−0.711 616 0	0.727 046 8	1.438 662 8	−0.159 442 00	0.004 433 800	−0.004 665	0.154 351 575
40	0.153 872 1	−0.449 122 0	0.462 963 4	0.912 085 4	−0.107 914 25	0.007 544 000	−0.000 989	0.098 734 075
80	0.113 721	−0.321 685 0	0.384 818 2	0.706 503 2	−0.077 509 50	−0.000 134 150	0.000 292 6	0.076 961 525
160	0.081 559 1	−0.257 724 0	0.269 273 5	0.526 997 5	−0.057 190 25	−0.001 642 500	−0.001 815	0.048 942 200
500	0.045 105 7	−0.158 407 0	0.198 534 1	0.356 941 1	−0.029 016 75	0.001 465 750	0.001 655 1	0.031 811 050
1 000	0.030 490 1	−0.078 449 0	0.094 403 1	0.172 852 1	−0.022 423 50	0.000 527 350	0.000 091 2	0.021 553 250

——Peter Y.Chen and Autumn D.Krauss
（马妍译　赵锋校）

参考文献

Chen, P. Y., & Popovich, P. M. (2002). *Correlation: Parametric and nonparametric measures*. Thousand Oaks, CA: Sage.

Hays, W. L. (1994). *Statistics* (5th ed.). New York: Harcourt Brace.

稳健的（Robust）

在统计估计中，稳健意味着对一个估计量【Estimator】取得最佳值的理想假定的小偏离的不敏感。“小”一词既可能表示对所有或几乎所有的数据点的小偏离，也可以表示对一小部分数据点的大偏离。

——Tim Futing Liao
（马妍译　赵锋校）

稳健标准误（Robust Standard Errors）

统计推论是以变量的观测值及关于变量的潜在分布和变量间关系的假设为基础的。如果一个模型对于这些假设【Assumption】的小偏离相对不敏感，那么它的参数估计【Parameter Estimation】和标准误【Standard Error】都被认为是稳健的。在某些情况下，对于某个假设的背离只会影响标准误。例如，对最小二乘【Ordinary Least Squares，OLS】的恒定误差方差假设和独立假设的违背虽然会得出无偏的【Unbiased】参数估计，但也会得出不一致的标准误。换句话说，通常的 OLS 的标准误是不可靠的，而这又会引发对统计显著性检验的疑问。克服这个问题的一种方法是使用稳健标准误。

稳健标准误的最初发展是为了补偿一种模式未知，又无法根本剔除的异方差性【Heteroskedasticity】的存在。存在着几种方法，可以计算一致的异方差性（heteroskedasticity-consistent）标准误；它们被分别称为 White，Eicker 或 Huber 标准误。这些标准误之大部分都是怀特（White，1980）提出的最初方法的变式，而怀特的方法使用的是一致的异方差性协方差矩阵（heteroskedasticity-consistent covariance matrix，HCCM）。HCCM 矩阵估计量是：

$$V(b)=(X'X)^{-1}X'\hat{\Phi}X(X'X)^{-1}$$

其中 $\hat{\Phi}=\text{diag}[e_i^2]$，$e_i$是 OLS 残差。

因为 HCCM 被找出并不是根据一个形式的异方差性模型，而是仅仅用源自 OLS 的回归项和残差作它的计算，所以怀特修正法可以很容易地被改写，以适于许多不同的目的。例如，针对来自整群抽样的数据，稳健标准误也可以用于放松独立假设。此外，稳健标准误（比如 Newey-West 标准误）也可以用来增进带有自相关【Autocorrelated】误差的合并的时间序列【Time-Series】数据和定群【Panel】数据的统计推论。

这里必须提及关于稳健标准误使用的

一些注意事项。首先,就小的样本量而言,不是所有的方差的稳健估计量都能有一致的上佳表现(详细讨论参见 Long & Ervin, 2000)。也许更重要的是,稳健标准误不应被视为审慎的模型设定的替代品。特别是如果异方差性的模式是已知的,那么运用加权最小二乘【Weighted Least Squares】通常可以更有效地对其作出修正,而且估得的模型也会更有效。同样,尽管稳健标准误能用于整群数据,但有关此类数据中存在的模式的更多信息——如可归之于群集效应的方差,可用多层次【Multilevel】模型求得。

最后,带有稳健标准误的回归【Regression】不应与"稳健回归"(robust regression)相混淆,后者是指回归模型能够承受偏斜分布和有影响的个案的干扰。使用此类稳健回归模型,所得出的估计值和标准误不同于那些由 OLS 回归得出的估计值和标准误。

——Robert Anderson

(马妍译 赵锋校)

参考文献

Huber, P. (1981). *Robust statistics*. New York: Wiley.

Long, J. S., & Ervin, L. H. (2000). Using heteroskedasticity consistent standard errors in the linear regression model. *American Statistician*, *54*, 217-224.

Newey, W. K., & West, K. D. (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*, 703-708.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*, 817-838.

角色扮演(Role Playing)

角色扮演是研究人际间行为的研究者常用的技术。使用该技术时,研究者给研究参与者分配特定的角色,指示参与者把一组假定的条件当作真实的情境,并采取行动。

在社会科学研究中,角色扮演的应用已有很长的历史。库尔特·卢因(Lewin,1939/1997),斯坦利·米尔格拉(Milgram,1963)和菲利普·津巴多在一些早期的经典的社会心理学实验中就使用了(Haney, Banks, & Zimbardo, 1973)。例如,津巴多和同事在斯坦福心理学系的地下室设立了一个模拟监狱,随机地指派大学生志愿者去扮演犯人或看守的角色。为了增加角色任务的真实性,那些被安排做犯人的人在寝室或课堂上被"逮捕",被要求穿上犯人的制服,被关进模拟监狱的临时"牢房"里。那些被指派为看守的人,穿着平时穿的衣服,还被分配了道具,如太阳镜、哨子和警棍。这一角色分配的反应是戏剧性的。那些被指派为看守的人的行为变得更加专横和更有攻击性,而那些被指派为犯人的人的行为变得更被动和更顺从。最终,当几个模拟的犯人开始出现非常痛苦的心理症状时,研究者提前终止了研究。在一个更近期的例子中,约翰逊(Johnson, 1994)利用一个角色扮演实验显示出,被指派为一个模拟组织的管理者表现出不同于被指派为下属的对话行为。

在一项社会研究中,角色扮演的过程可以被比作一项计算机模拟【Computer Simulation】,其中人的心灵像处理器一样工作,而不像一台计算机。这些技术将人类能力当作资本,将之投入创造性的角色扮演中——去实现某些思想规则、感觉和行为。在一定的既定条件下,人类也许能有效地扮

演任意指派的角色得到了戈夫曼(Goffman,1959)的观点——所有社会行为都是一种戏剧表演形式的支持。在研究中应用角色扮演假定,个人有能力有效地想象在一组给定的环境下,他们可能如何行动。他还假定,个体在模拟的环境下如何感知和行动,同他们在实际环境(他们正模拟的)中可能如何感知和行动之间存在着一种关系。然而,角色扮演在社会科学研究中的效度【Validity】并不依赖于个人的(表演)能力,即准确地按照他们可能在被模拟的真实环境下行为的方式去做。毋宁,大多数利用角色分配的研究的效力在于个体间行为的比较,而这些个体被随机分配去占据不同角色。因此,更多情况下,在占据不同角色的个体之间的行为、思想和感觉中,观察到的差异比基于角色的行为的实际特征更有意思。

——Dawn T. Robinson

(王玥译　赵锋校)

参考文献

Goffman, E. (1959). *Presentation of self in everyday life*. Garden City, NY: Doubleday.

Haney, C., Banks, C., & Zimbardo, P. (1973). Interpersonal dynamics in a simulated prison. *International Journal of Criminology and Penology*, *1*, 69-97.

Johnson, C. (1994). Gender, legitimate authority, and leadersubordinate conversations. *American Sociological Review*, *59*, 122-135.

Lewin, K. (1997). Experiments in social space. In G. W. Lewin (Ed.), *Resolving social conflicts: Field theory in social science* (pp. 59-67). Washington, DC: American Psychological Association. (Originally published in 1939)

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, *67*, 371-378.

均方根(Root Mean Square)

均方根误差(root mean square error,RMSE)是衡量某个估计量估计程度的一个量度。如果$\hat{\beta}$是β的一个估计量,那么$\hat{\beta}$的RMSE可以表示为:

$$\text{RMSE}(\hat{\beta}) = E[(\hat{\beta}-\beta)^2] \qquad (1)$$

我们会注意到,如果$\beta = E[\hat{\beta}]$,那么它就等于$\hat{\beta}$的方差。将式(1)进行改写,并对等式的右边进行简化,可得到如下的分解:

$$\begin{aligned}\text{RMSE}(\hat{\beta}) &= E[(\hat{\beta}-E[\hat{\beta}]+E[\hat{\beta}]-\beta)^2] \\ &= E[(\hat{\beta}-E[\hat{\beta}])^2+(E[\hat{\beta}]-\beta)^2-2(\hat{\beta}-E[\hat{\beta}])(E[\hat{\beta}]-\beta)] \\ &= E[(\hat{\beta}-E[\hat{\beta}])^2]+E[(E[\hat{\beta}]-\beta)^2]-2E[(\hat{\beta}-E[\hat{\beta}])(E[\hat{\beta}]-\beta)] \qquad (2)\end{aligned}$$

最后一项表达式趋于0,最后有:

$$\text{RMSE}(\hat{\beta}) = \text{Var}(\hat{\beta}) + [\text{bias}(\hat{\beta})]^2 \qquad (3)$$

再次注意，如果 $\hat{\beta}$ 是 β 的无偏估计量，那么 $\mathrm{RMSE}(\hat{\beta}) = \mathrm{Var}(\hat{\beta})$。

这就导致了对两个估计量的根本比较：即有两个 β 的估计量 $\hat{\beta}$ 和 $\hat{\beta}^*$，其中，$\hat{\beta}$ 是 β 的无偏【Unbiased】估计量，而 $\hat{\beta}^*$ 是 β 的有偏【Biased】估计量，但是 $\hat{\beta}^*$ 的方差比 $\hat{\beta}$ 的方差小。在这种情形中，根据两个估计量的相对方差，以及 $\hat{\beta}^*$ 的偏倚程度，$\hat{\beta}^*$ 的 RMSE 可能比 $\hat{\beta}$ 的要小。这时，要想极小化 RMSE，就可能偏好有偏的估计量，而不是无偏的估计量。

RMSE 和 OLS

RMSE 还有另一个常见用法。有些统计软件包在给出最小二乘【Ordinary Least Squares，OLS】回归结果时，还会列出 RMSE 的值。这个值就是最常说的回归标准误：

$$\mathrm{RMSE} = \sqrt{\frac{\sum(\hat{Y} - Y)^2}{n - k}} \tag{4}$$

这是总体【Population】扰动项的标准差【Standard Deviation】的一个估计量。或者，如果我们有一个典型的 OLS 设定，如：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \tag{5}$$

那么，RMSE 是对 σ 的估计，其中 σ^2 是 ε 的方差。

因此，在用方程的估计值预测时，RMSE 提供了预测所产生的误差的成分信息：它是对误差的随机【Stochastic】构成的一个量度。在知道模型的真实参数【Parameter】的前提下，它给我们提供了一个用以衡量模型对数据生成过程（data-generating process）的拟合度的量度。RMSE 的较大值意味着因变量上有大量的方差未被模型所解释。而且它意味着，即使我们获得了对模型参数非常精确的估计，我们所做的任何预测【Prediction】根据 RMSE 仍将有很大程度的不确定性。

RMSE 是使用模型【Model】估计值进行预测产生的任意点预测值误差的一个组成部分。就某个点预测值的不确定性【Uncertainty】考察还需要考虑所使用的参数估计值的估计误差。因此，在计算任意点估计 $\hat{Y}_i$ 的不确定性时，进入 Y_i 预测中并存在于估计值 $\hat{\beta}$ 上的不确定性可能会缩小包含在 RMSE 中的随机不确定性。

RMSE 也是一个拟合优度【Goodness-of-Fit】的量度。但它是对模型的拟合优度的量度，而不是估计值的量度。它通常被拿来和另一个拟合优度量度 R 方【R-Squared】进行比较。然而，R^2 包含了与 RMSE 不同的信息。RMSE 是方差的随机构成的一个估计量，而 R^2 是对样本中 Y 的预测值和估计值间的拟合的一个量度。从数学上讲，两者的关联如下：

$$\mathrm{RMSE}^2 = \mathrm{Var}(Y) - \mathrm{Var}(\hat{Y}) \tag{6}$$

$$R^2 = \frac{\mathrm{Var}(\hat{Y})}{\mathrm{Var}(Y)} \tag{7}$$

所以，如果想知道我们对给定样本的数据拟合得怎样，R^2 就可能是一个有效的量度。但是，如果想了解出现的随机方差的量，就需考察 RMSE。R^2 会随着不同的样本中 Y 的方差而变化，而 RMSE 本身是不同样本中随机方差的一个一致的估计量。RMSE 和 R^2 之间的差异在阿肯（Achen，1977）和金（King，1990）的论文中有更详细的描述。

——Jonathan Nagler

（马妍译　赵锋校）

参考文献

Achen, C. H. (1977). Measuring representation: Perils of the correlation coefficient. *American Journal of Political Science*, *21*, 805-815.

King, G. (1990). Stochastic variation: A comment on Lewis-Beck and Skalaban's "The R-Squared". *Political Analysis*, *2*, 185-200.

旋转因子(Rotated Factor)

参见因子分析【Factor Analysis】。

旋转(Rotations)

因子分析【Factor Analysis】的不同方法首先都需要从一个数据集中萃取一组因子。这些因子几乎总是正交的,并且按照它们解释掉原始数据的方差【Variance】的比率依次排列。一般而言,仅保留下一个所有因子的(小的)子集用于进一步的考察,而剩余的因子要么被当作无关的,要么被当作不存在的(即假定它们反映的只是测量误差或测量噪声)。

为了解释那些据认为是相关的因子,通常,在第一次选取步骤后,随之要进行对所保留下的因子的旋转。常用的两个主要旋转类型是正交的(orthogonal)(即新的轴之间也彼此是正交的)和斜交的(oblique)(即新的轴彼此之间不需要是正交的)。由于旋转总是在一个子空间进行(即所谓的因子空间),新的轴将比原来的因子(根据最佳原则计算得到)解释掉较少的方差,但是,在旋转后,总的子空间解释掉的方差部分同旋转前的部分相等(变化的只是方差的分割方式)。由于旋转后的轴不再根据某一统计准则来确定,因此它们的存在理由即在于方便解释。

在本词条中,用经主成分分析【Principal Components Analysis,PCA】(也称为 R 模式)得到的变量的负荷作为旋转程序的示例说明,不过,此处描述的方法对其他类型的分析,以及对受试者得分(所谓的 Q 模式)的分析,同样是有效的。

在继续介绍之前,要特别强调:因为旋转总是发生在一个子空间中(即被留下的因子的空间),所以该子空间的旋转会强烈地影响旋转的结果。因此,在因子分析的旋转实践中,强烈建议尝试用若干规模不等的余留因子的子空间,以便评估旋转解释的稳健性。

记号法

为了使讨论更具体,先要介绍主成分分析框架内的旋转。主成分分析始于一个数据矩阵,记作 $\mathbf{Y}$,该矩阵有 I 行 J 列,其中每一行代表用 J 个测量描述的一个单元(一般而言,即一个受试者),而 J 个测量常表示作 z 分数。随后这个数据矩阵被分解为(受试者的)得分或成分和变量的负荷(通常,负荷是原始变量和由分析所萃取的成分之间的相关系数)。形式上,主成分分析等价于数据矩阵的奇异值分解,如下:

$$\mathbf{Y}=\mathbf{P}\Delta\mathbf{Q}^{T} \quad (1)$$

有约束条件 $\mathbf{Q}^{T}\mathbf{Q}=\mathbf{P}^{T}\mathbf{P}=\mathbf{I}$(其中 $\mathbf{I}$ 是单

位矩阵),其中 $\mathbf{\Delta}$ 是在对角线上具有所谓奇异值的对角矩阵(diagonal matrix)。正交矩阵 $\mathbf{Q}$ 是一个负荷矩阵(或原始变量对成分的预测),其中一行代表一个原始变量,而一列代表一个新的因子。通常,有受试者得分矩阵 $\mathbf{F}=\mathbf{P\Delta}$(有的作者用 $\mathbf{P}$ 代表得分)。

旋转:何时和为何?

因子旋转的大部分原理来自瑟斯通(Thurstone,1947)和卡特利(Cattell,1978),而他们的辩护理由是:因为该程序可以简化因子结构,进而使得因子结构的解释更容易、更可信(即易于用不同的数据样本进行重复)。

瑟斯通提出了可以识别一个简单结构的五项准则。根据这些仍然在文献中报告的准则,当下述条件满足时,一个负荷矩阵(其中行对应于原始变量,而列对应于因子)是简单的。具体包括:

①每一行至少包含一个零。

②就每一列而言,至少有与列数(即余留的因子数量)同样多的零。

③就任意一对因子而言,有些变量在一个因子上有零负荷,而在另一个因子上则有大的负荷。

④就任意一对因子而言,存在着相当大比率的零负荷。

⑤就任意一对因子而言,仅存在着少量的大的负荷。

因子旋转可以(也常常)以作图的方式完成,但是大部分通过数学分析完成,为了用计算机软件包完成它,还必须有简单结构的数学设定。

正交旋转

一个正交旋转可以设定为一个旋转矩阵,记作 $\mathbf{R}$,其中行代表原来的因子,列代表新的(旋转的)因子。在 m 行和 n 列的交叉处,有一个原始轴和新轴的夹角的余弦函数:$r_{m,n}=\cos\theta_{m,n}$。例如,例示在图 1 中的旋转可由下列矩阵予以刻画:

$$\mathbf{R}=\begin{bmatrix}\cos\theta_{1,1} & \cos\theta_{1,2}\\ \cos\theta_{2,1} & \cos\theta_{2,2}\end{bmatrix}=\begin{bmatrix}\cos\theta_{1,1} & -\sin\theta_{1,1}\\ \sin\theta_{1,1} & \cos\theta_{1,1}\end{bmatrix}\quad(2)$$

有 $\theta_{1,1}=15°$。一个旋转矩阵具有正交的重要属性,因为它对应于一个方向余弦函数矩阵,因此有 $\mathbf{R}^T\mathbf{R}=\mathbf{I}$。

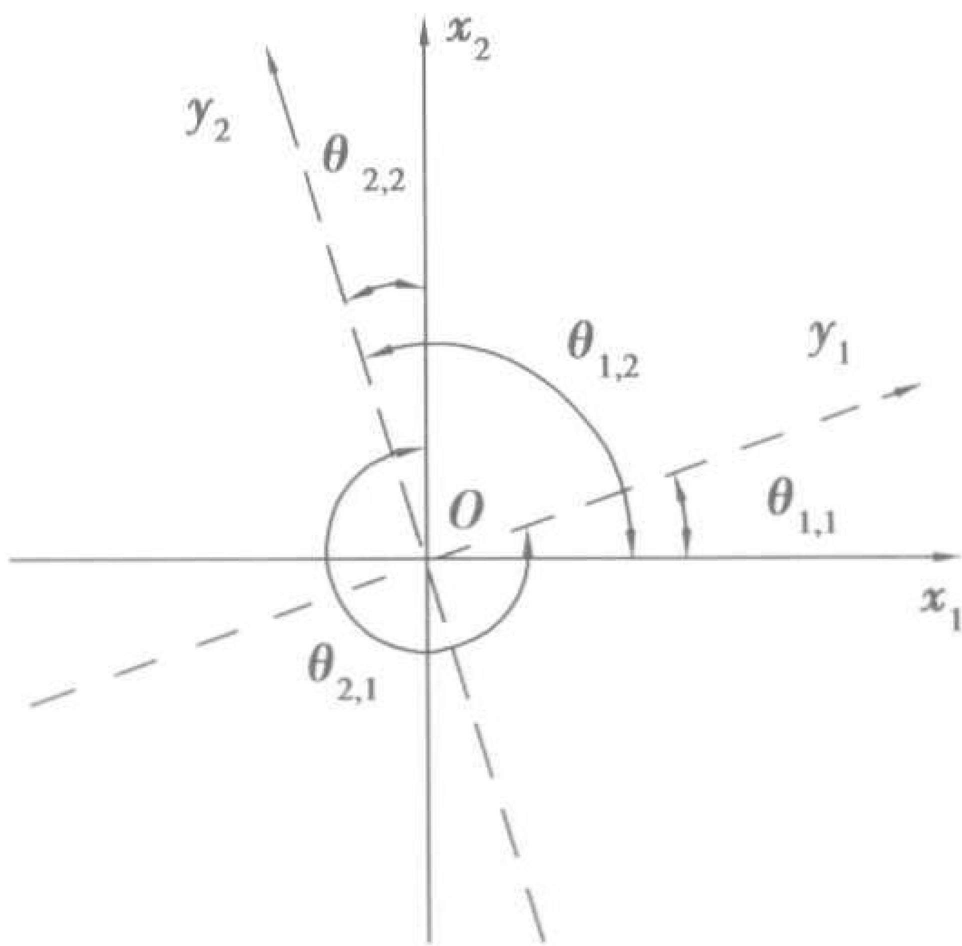


图 1 在二维空间中一个正交旋转;一个原始轴 m 和一个新轴 n 之间的旋转角度记作 $\theta_{m,n}$

方差最大法

由凯泽(Kaiser,1958)发展出的方差最大法(VARIMAX),无疑是迄今为止最通行的旋转方法。就方差最大法而言,一个简单的解意味着每一个因子具有一组数量少的大负荷和数量多的 0(或小)负荷。这就简化了因子解释,因为在一次方差最大法的旋转后,每一个原初的变量倾向于同一个(或数量较少)的因子相关,同时,每一个因子只代表数量较少的一组变量。此外,因子常可以根据少数带有正负荷的变量同少数带有负负荷的变量的对立,而予以解释。

形式上,方差最大法即找到一个原始因子的旋转(即原始因子的一个线性组合),从而使得负荷的方差是极大的,这相当于极

大化，

$$V = \sum (q_{j,l}^2 - \bar{q}_{j,l}^2) \quad (3)$$

$q_{j,l}^2$是第 j 个变量在第 l 个因子上的负荷的平方，而 $\bar{q}_{j,l}^2$是负荷平方的均值（为了计算的稳定性，在最大化程序前，负荷矩阵中的每个负荷通常以长度 1 为标准）。

其他正交旋转法

还有其他几种正交旋转方法，比如四方最大旋转（quartimax rotation），该方法极小化了用于解释每个变量所需的因子数；以及均等最大旋转（equimax rotation），该方法是方差最大法和四方最大法的一个折中。还有其他的方法，但是没有一种方法像方差最大法这样普及。

正交方差最大旋转的一个例子

为了用示例说明方差最大旋转的程序，设想有五种葡萄酒，以及一组专家对于它们的享受度、某种酒多适合配甜点来享用，以及某种酒多适合配肉来享用的平均评定分；相关的描述还包括每瓶酒的价格、糖度、酒精度和酸度。数据都显示在表 1 中。

表 1　一个关于 PCA 和旋转的（虚构）例子；以 7 个变量描述的 5 种葡萄酒

	享受度	配肉	配甜点	价格	糖度	酒精度	酸度
葡萄酒 1	14	7	8	7	7	13	7
葡萄酒 2	10	7	6	4	3	14	7
葡萄酒 3	8	5	5	10	5	12	5
葡萄酒 4	2	4	7	16	7	11	3
葡萄酒 5	6	2	4	13	3	10	3

表 1 中的 PCA 提取了 4 个因子（特征值分别是 4.762 7，1.810 1，0.352 7 和 0.074 4），其后，解释了 94% 的方差的一个两因子的解（对应于特征值大于 1 的成分）被保留下来用于旋转。表 2 给出了两因子解的负荷。

表 2　葡萄酒案例：7 个变量对最初的两个成分的原始负荷

	享受度	配肉	配甜点	价格	糖度	酒精度	酸度
因子 1	−0.396 5	−0.445 4	−0.264 6	0.416 0	−0.048 5	−0.438 5	−0.454 7
因子 2	0.114 9	−0.109 0	−0.585 4	−0.311 1	−0.724 5	0.055 5	0.086 5

用于负荷表的方差最大旋转程序给出了一个 15°的顺时针旋转（其对应的余弦值为 0.97）。见表 3，这就生成了一组新的旋转因子。旋转程序示例于图 2、图 3 和图 4 上。在该例子中，解释的简洁性的改进较为有限，这可能是因为如此小的数据集的因子结构已经非常简单了。第一个维度仍然与价格有关，而第二个维度作为甜度的维度现在看上去更清晰了。

表 3　葡萄酒案例:方差最大法旋转后,7 个变量对最初的两个成分的负荷

	享受度	配肉	配甜点	价格	糖度	酒精度	酸度
因子 1	−0.412 5	−0.405 7	−0.114 7	0.479 0	0.128 6	−0.438 9	−0.462 0
因子 2	0.015 3	−0.213 8	−0.632 1	−0.201 0	−0.714 6	−0.052 5	−0.026 4

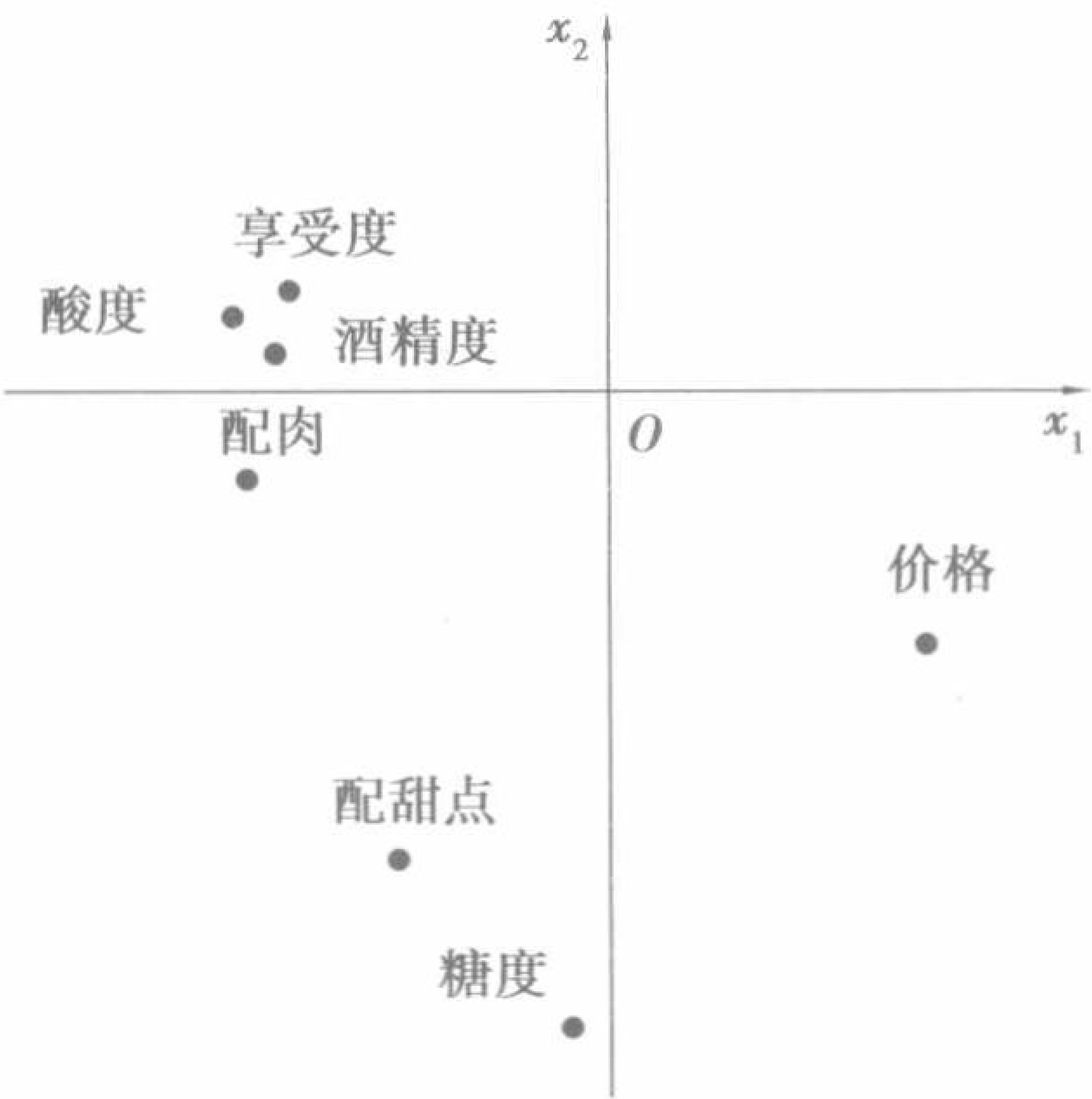


图 2　葡萄酒案例中 7 个变量的原始负荷,PCA

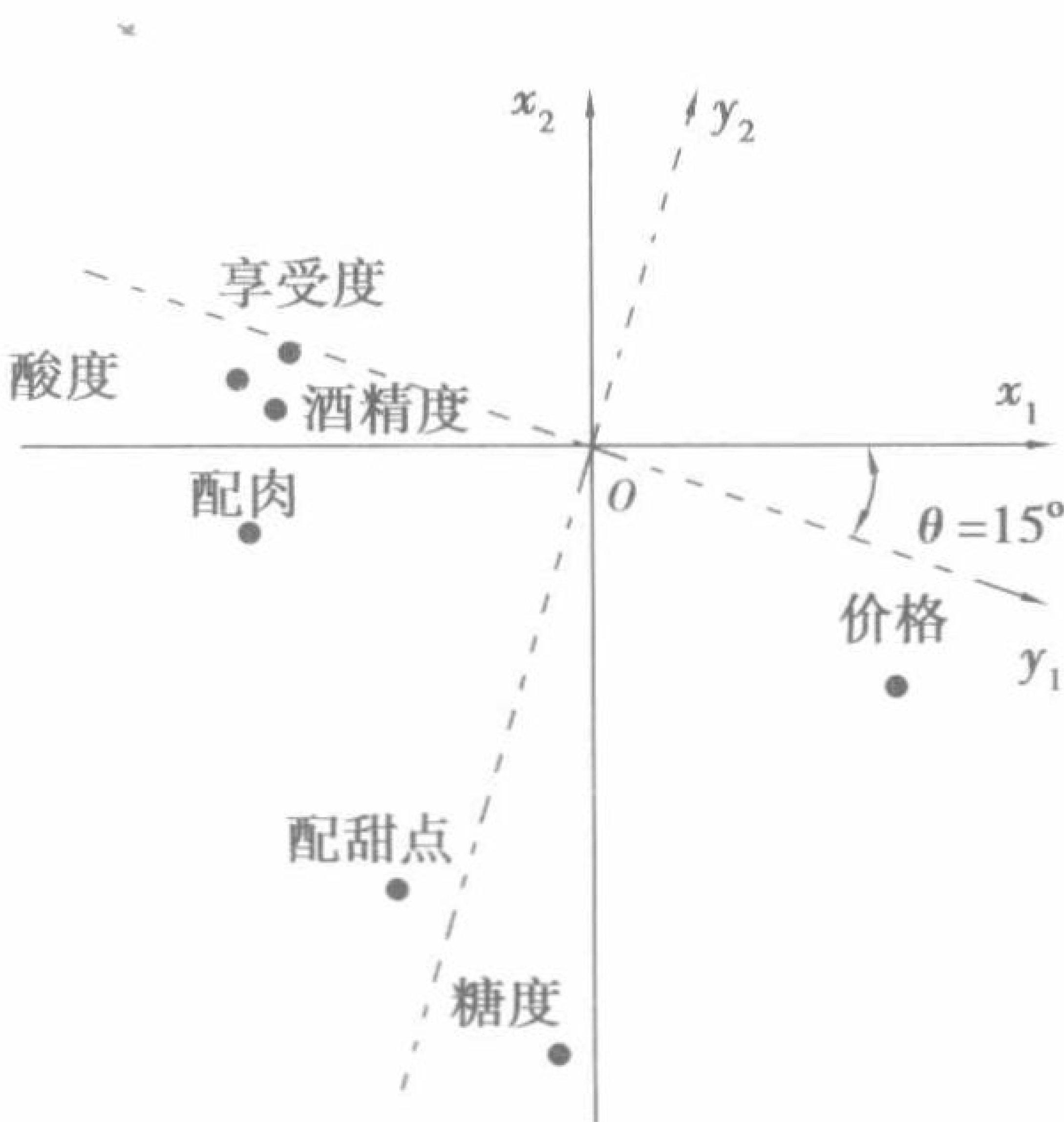


图 3　葡萄酒案例中 7 个变量的负荷,原始轴和得自于方差最大法的新(旋转)轴,PCA

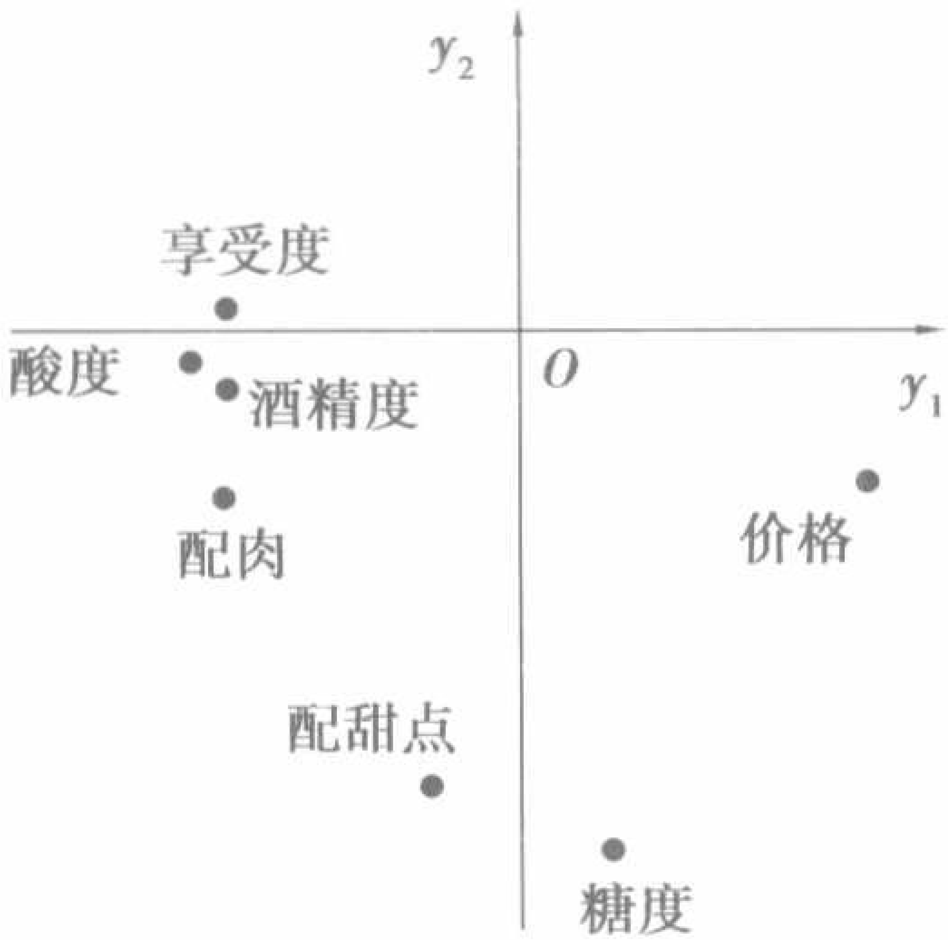


图 4　葡萄酒案例中 7 个变量在方差最大法旋转后的负荷,PCA

斜交旋转

在斜交旋转中,新轴可以自由地选取因子空间中的任何位置,但是,由于两个高度相关的因子最好被解释成唯一的因子,因此因子间容许的相关程度一般都很小。因为,为了获得解释上的简洁性,斜交旋转放松了正交限定(orthogonality constraint)。虽然瑟斯通强烈推荐这种方法,但是与它们的竞争者正交旋转相比,斜交旋转更少被使用。

对于斜交旋转,promax 旋转的优势是快速并且概念上简单。其名字来源于强制旋转(procrustean rotation),因为它试图拟合一个具有简单结构的目标(target)矩阵。promax 旋转包含两个必要的步骤:第一步是界定目标矩阵,而该矩阵几乎总是得自于一个方差最大旋转的结果,即将方差最大旋转的解幂乘(通常介于 2~4),以此“迫使”目标矩阵的负荷结构成为两级的。第二步的结果则得自于计算从方差最大解到目标矩阵的一个最小二乘拟合。

即使原则上斜交旋转的结果可以用图形呈现,它们也几乎总是通过检查旋转后的轴同原始变量间的相关来解释。这些相关被解释成因子负荷。就葡萄酒案例而言,首先将方差最大解幂乘以 4,从而得到目标矩

阵,而 promax 旋转得到的负荷见表 4。从这些负荷中可以发现:第一个维度再次分离出了价格(同因子正相关),同时对立于其他变量——酒精度、酸度、配肉及享受度。而第二个轴显示出它与第一个轴只有非常微弱的相关(r =−0.02)。它对应于"甜美\配甜品"因子。再次,这一解释与原来的 PCA 很相似,因为小规模数据集本身就倾向有结构简单的解。

表 4 葡萄酒案例:promax 斜交旋转解;7 个变量对最初的两个成分的相关

	享受度	配肉	配甜点	价格	糖度	酒精度	酸度
因子 1	−0.878 3	−0.943 3	−0.465 3	0.956 2	0.027 1	−0.958 3	−0.998 8
因子 2	−0.155 9	−0.475 6	−0.939 3	−0.076 8	−0.950 7	−0.262 8	−0.236 0

斜交旋转不及它的竞争对手正交旋转受欢迎,但是这个趋势可能因为新技术的出现而改变,如独立成分分析(independent component analysis)的发展。这一新近的技术最初创发自信号处理和神经网络【Neural Networks】领域,可直接从数据矩阵中求出一个使得统计独立性极大化的斜交解。

——Hervé Abdi

(马妍译 赵锋校)

参考文献

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis*. New York: Wiley.

Kim, J. O., & Mueller, C. W. (1978). *Factor analysis: Statistical methods and practical issues*. Beverly Hills, CA: Sage.

Kline, P. (1994). *An easy guide to factor analysis*. Thousand Oaks, CA: Sage.

Mulaik, S. A. (1972). *The foundations of factor analysis*. New York: McGraw-Hill.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Reyment, R. A., & Jöreskog, K. G. (1993). *Applied factor analysis in the natural sciences*. Cambridge, UK: Cambridge University Press.

Rummel, R. J. (1970). *Applied factor analysis*. Evanston, IL: Northwestern University Press.

取整误差(Rounding Error)

人们通常习惯于对估计量进行四舍五入。如果一个估计量被四舍五入,那么它的位数会减少。例如,一个收入估计量为 67 542.67美元,它会被四舍五入为 67 543 美元,因此,减少了两位小数。我们注意到收入被向上入了一位,末尾的 67 美分被入为 1 美元。当然如果最初的估计量是 67 542.38美元,我们会将其下舍为 67 542 美元。在所有四舍五入的例子中,无论是上入或是下舍,都产生了一些误差。在第一个例子中,误差是 0.67;在第二个例子中,误差是 0.38。取整误差出现的一个常见情形是,当我们加总一张表中的列百分比时,结果经常是 99% 或者 101%,而非正确的总和 100%。当出现这种情况时,分析者通常在脚注中说明比例总和不等于 100%是由于取

整误差。取整误差的存在是常有的事,并非不可避免,而且很少造成麻烦,因为误差很小并且是随机的。当然,也有可能出现取整过多的情况。例如,假定将前文例子中的收入取整为 70 000 美元,那么收入估计量中包含的取整误差几乎是 2 500 美元。取整误差因此增加了,但这是一个很大的误差吗?这个问题的答案取决于所要回答的研究问题【Research Question】和分析者想要达到的精度。在另一种极端的情况下,也可能出现取整不充分的情况。例如,在大部分收入研究中,我们无须知道以美分为测量单位的收入。事实上,那种声称能够将收入估计到美分的说法,可能造成一种虚假的精确感,而否认了测量【Measurement】中固有的更大误差。

——Michael S. Lewis-Beck

(马妍译　赵锋校)

行关联(Row Association)

参见关联模型【Association Model】。

行-列关联(Row-Column Association)

参见关联模型【Association Model】。

R 方(*R*-Squared)

R 方衡量了一个回归模型的解释力或预测力。它是一个拟合优度【Goodness-of-Fit】量度,表明线性回归方程拟合数据的好坏。

在回归分析中,评估估得的回归方程的表现非常重要。它在多大程度上说明了所研究的现象?对一个简单的或多元回归【Multiple Regression】模型而言,*R* 方是首要的方程表现的量度。设想一个政策分析者正在研究美国 50 个州公立学校的支出。分析者设定了一个简单的回归模型,$Y = a + bX + e$,其中 Y 是因变量【Dependent Variable】——2000 年,每个州每一名学生的公立学校支出(千美元),X 是自变量【Independent Variable】——城市化水平(根据 2000 年人口普查,人口数大于 25 000 的城市中城市人口的百分比),而 e 是误差项。该模型认为,公立学校的花费可以部分地由城市化水平来解释。假设用最小二乘【Ordinary Least Squares, OLS】估计得到如下结果:$Y= 2.11 + 0.60X+e$,这就表明了城市人口每增加百分之一,一个州的期望支出会增加 600 美元。这个方程的 *R* 方或者确定系数是0.42,它表示城市化水平解释了或者至少预测了州公立学校支出中 42%的方差。

R 方给出了,同未知 X 相比,已知 X 时,预测 Y 所得精度的增量。设若分析者已知 Y 的取值,但并不知道这些取值属于哪些州,还想预测每个州的学校支出。最佳推测(即极小化误差)就是 Y 的均值,即 Y_m。这个推测与绝大部分州有很大差异,亦即观测

的支出值与平均的支出值之间($Y-Y_m$)有很大差距。将所有这些距离合在一起(在取平方以排除加号和减号互相抵消的影响之后),它们就代表了未知 X 时的总预测误差,亦即总离差平方和(total sum of squared deviations,TSS) $TSS=\sum(Y-Y_m)^2$。

回归分析承诺,对 X 的知识和它对 Y 的线性关系可以减少这一预测误差。例如,已知一个州的 X 取值是30%的城市化率,Y 的预测值 $\hat{Y}$ 就是20.12($\hat{Y}=2.11+18.10=20.21$),而不是 Y_m。Y 的预测值与 Y 的均值之间的距离($\hat{Y}-Y_m$)是由于已知 X 而得到的超过基准预测的增量。这一距离(对所有观测值,先平方后求和)是可归于应用回归线而得到的预测误差的减量;此即,回归平方和(regression sum of squared,RSS)$RSS=\sum(\hat{Y}-Y_m)^2$。除非回归线对每个个案都有完美的预测,否则总会遗留一些误差:Y 的观测值与预测值之间的距离。这些距离(先平方再加总)代表了变量 Y 上未被解释掉的方差;此即,误差平方和(error sum of squared,ESS)$ESS=\sum(Y-\hat{Y})^2$。

于是,因变量上的总方差TSS有两个独一无二的部分:由回归解释掉的RSS和未被回归解释掉的ESS。R 方反映了回归掉的部分同总误差的比率,即 $RSS/TSS=R$ 方。R 方的统计范围是0~1;当所有的方差都被解释掉时为1,当没有方差被解释掉时为0。就现实世界的数据而言,R 方很少达到这些极限值。在前文的例子中,可以说公立学校支出上42%的方差被城市化水平解释了。重要的是,这个解释可能更多地属于“统计的”解释而非“因果的”。可能是城市化水平有助于预测公立学校的支出,但没有在一定的理论意义上解释它。在这种情况下,更谨慎的,同时也许更正确的说法是,X 只是“说明了”变量 Y 上如此多的方差。在双变量回归的例子中,R 方等于相关【Correlation】系数的平方(r^2)。

在一个多元回归方程中,R 方被称作多重确定系数。它衡量了模型的线性拟合优度,其范围也是从1到0。例如,当我们用OLS估计多元回归模型 $Y=a+bX+cZ+e$ 时,R 方表示一个平面对一个三维空间中的点的拟合程度。当自变量超过2个时,则会拟合一个超平面。多重相关【Multiple Correlation】系数,记作 R,是 R 方的平方根,表示全体自变量和因变量的相关系数。

如果一个自变量被加入一个回归模型中,R 方总是会增加。增量部分可归于机会,即由于增加自变量使用了自由度【Degrees of Freedom】。特别是随着自变量的数量不断接近样本量,R 方的估计将会夸大模型对数据的真实拟合。对于多元回归模型,绝大多数分析者在报告 R 方的同时也报告调整的 R 方【Adjusted *R*-Squared】,或者以后者取代前者。调整的 R 方在数值上小于 R 方,但不会小太多。与 R 方不同,随着新的自变量被加入模型,调整的 R 方可能下降(极少的情况下,如果未调整的拟合度极低时,调整的 R 方可能是负的。而未调整的 R 方只有当常数项被迫为0时,才有可能是负的)。

分析者通常偏好一个较高的 R 方,认为一个较好的拟合意味着一个更佳的解释。然而,这可能是真的,也可能不是。R 方的最大化不应当成目的本身。将那些同因变量有高相关的变量盲目地加入回归模型的变量之中,这将产生一个高的 R 方值。但是,根据一个后来的样本估计该模型时,原本高的拟合很可能骤然下降,因为原本高的 R 方中的大部分拟合来自第一个样本中的机会方差。一个适度的甚至是较低的 R 方并不必然不好。某些类型的数据,如民意测验,一般不会有较高的拟合。不管是何种数据,一个低的拟合只是意味着只有小部分的

Y 能被解释，其余部分是随机的。有一种情况，低的 *R* 方可能是一个不好的标志，即非线性【Nonlinearity】的情况。如果两变量间的关系更接近一条曲线而非一条直线，那么线性设定和用 *R* 方评估模型的拟合表现都是不恰当的。最后，对所有模型而言，一个低的 *R* 方表明模型有较差的预测。

R 方备受批评。有人认为，用于预测比较的基线 Y_m 太缺乏可比性。然而，通常很难提出一个更好的替代基线。同样，有人提出异议，认为 *R* 方只能评估线性模型的拟合程度。可是在实践中，通常很难改进线性设定。进一步的问题是，不同模型的 *R* 方是否能检验它们之间差异的统计显著性，以便评估竞争的模型设定。这点仅当一个模型是另一个较大模型的嵌套子模型时才能实现。合理地运用这类显著性检验部分取决于 *R* 方是被当作一个描述统计量，还是一个推断统计量。少数分析者完全拒绝 *R* 方，而喜欢用 *Y* 的估计标准误【Standard Error of Estimate，SEE】作为拟合优度的量度。单独使用 SEE 量度的困难在于，它不具有传统的理论边界的上限（如 1.0）。随着 SEE 变大，人们知道存在着更多的预测误差，但是人们并不知道它是否已经达到了它的最大值。总之，应当同时报告这两个拟合度量度，因为它们反映了不同的东西。

——Michael S. Lewis-Beck

（马妍译　赵锋校）

＊也可参见确定系数【Coefficient of Determination】。

参考文献

Lewis-Beck, M. S., & Skalaban, A. (1990). The *R*-squared: Some straight talk. *Political Analysis*, 2, 153-170.

Pindyck, R. S., & Rubinfeld, D. L. (1991). *Econometric models and economic forecasts* (3rd ed.). New York: McGraw-Hill.

Weisberg, S. (1985). *Applied linear regression*. New York: Wiley.

S

样本(Sample)

样本是总体【Population】的要素或成员的一个子集。它用于收集来自或关于要素的信息,而所得信息可以代表被抽取的总体。抽样是一种有效的、节约成本的收集数据【Data】的方式,即它既可以降低成本,也能提升所得数据的质量。

降低成本源于仅对较少的分析单位收集信息。在典型的民意调查中,1 000 个样本量只代表美国公民的 1/288 400。这个容量的代表样本可以允许研究者以一定的精度估出美国人的平均年龄,同时,相对于对全部人口普查【Census】,也可以更迅捷地收集到信息。

使用样本而节省的经费还可用于提高数据质量。省得的经费可以通过减少无回答【Nonresponse】来提高实施原抽样设计的覆盖面(coverage)。多余的资源可用来做访员培训【Interviewer Training】以改进问卷【Questionnaire】调查的执行;也可用于增加问卷长度,对特定的概念提出更多问题,从而增加概念测量的信度【Reliability】和效度【Validity】。

每个样本都依据两个关键的属性得到评估,即它的设计和它的执行。为了调用抽样分布【Sampling Distribution】、应用置信区间【Confidence Interval】和置信水平,研究者必须使用某种概率设计,即总体的每个要素都有一个已知的、非零的被选中的机会。非概率设计,如配额样本【Quota Sample】、方便样本【Convenience Sample】或目的样本【Purposive Sampling】则不能使用上述统计技术。

每个样本设计必须得到有效的执行,投入资源去联系每一个要素,以便从中收集数据。覆盖率或回答率反映了:在样本中,已获得信息的要素数量同要素的原始数量的比率。即使使用概率设计,低的覆盖率也能使得样本估计的应用无效,因为要考虑信息的损失是系统的且非随机的。例如,在一个政府用的样本中,社会经济状况或住房状况常代表了误差【Error】的系统来源,因为低收入或无家可归的人可能难以找到。在公共舆论调查中,那些对政治不感兴趣的个体比那些对政治感兴趣的,更可能拒绝参与电话调查【Telephone Survey】,所以结果可能反映了那些态度强烈的人的情况,而不是总人口的情况。

——Michael W. Traugott

(王玥译　赵锋校)

参考文献

Kish, L. (1995). *Survey sampling*. New York: Wiley.

Lavrakas, P. (1993). *Telephone survey methods: Sampling, selection, and supervision*. Newbury Park, CA: Sage.

Levy, P. S., & Lemeshow, S. (1999). *Sampling of populations: Methods and applications*. New York: Wiley.

抽样(Sampling)

抽样的基本观念是从部分到总体的统计外推【Extrapolation】——从“样本”到“总体”[总体有时被神秘地称为“全体(universe)”]。一个最紧要的定理是:样本的选取必须能公正地代表总体。

选取样本的方法被称为“设计”。好的设计要使用概率方法,在选取调查单元的过程中要极小化主观判断。使用概率方法抽取的样本被称为“概率样本”。

在实际工作中,偏倚【Bias】是一项严重的问题;概率样本则可极小化偏差。然而,最终用于从概率样本外推到总体的方法应当将抽取样本的方法纳入考量之中;否则,偏差还是会从后门进来。

抽样的观念可用对人群的抽样或对商业记录的抽样来说明,不过它可以有更广泛的应用。比如,存在着对建筑物、农场、法律案件、学校、树木、当地工会及其他各种总体的抽样调查。

样本设计

应当区别概率样本与“方便样本”[也称为“偶遇样本(grab samples)”]。一个典型的方便样本可能是上研究者的导论课的学生。一个“购物中心样本”则由那些某几天中在特定购物中心购物并愿意参与访谈的人群组成。使用这一术语的原因很明显,它的含义也很明确:这种样本除了它自己以外不代表任何确定的更大总体。

为抽取一个概率样本,首先要界定目标总体。接下来要造出一个“抽样框”,即一个从中抽取样本单元的清单。一种简要的设计是简单随机抽样【Simple Random Sampling】。例如,抽取由100个单元构成的简单随机样本,可以从样本框中随机选出一个单元,将它放到样本中;然后,再从样本框中剩余的单元中随机地选出另一个单元,以此进行,直至选出100个单元为止。整个过程的每一步骤上,池中所有单元都有同样被选中的机会。

简单随机抽样经常把商业记录当作总体,即便这种总体的数量巨大。当它用于人时,特别是在开展面对面访谈的情况下,简单随机抽样很少切实可行:我们从哪里获得抽样框?因此需要更复杂的设计。例如,如果想要抽取城市中的人,可以列一张包含城市内的所有街区的清单,造出街区的样本框,从中抽取一个街区的简单随机样本,再访谈被选街区内所有住宅单元中的人。这就是一个“整群样本”,群即街区。

注意,必须十分谨慎地界定总体:它由那些抽样发生时城市中住宅单元内生活的人组成。还有一些变化形式,例如,可以只访谈家庭户中的一个人来获得整个家庭户的信息。或者,可以在家庭户内随机地选取一个人。可以限定回答人的年龄等特征。如果开展电话访问,通常随机数字拨号【Random Digit Dialing】可以作为简单随机抽样的合理近似——其目标总体是有电话的人。

误差的种类

因为样本只是总体的一部分,所以统计外推不可避免地引起误差【Error】。误差有两种:抽样误差(随机误差)和非抽样误差(系统误差)。非抽样误差也常被称为“偏差”,当然其中不含有偏见的意思。

抽样误差【Sampling Error】来自抽取一个样本时的抽取机会,结果是:我们得到的某一类单元太多,而另一类的则不足。抽样

误差可能的影响常使用“SE”或标准误【Standard Error】来加以量化。如果样本是概率样本【Probability Sampling】，标准误 SE 可以用样本设计和样本数据来估计。

随着“样本量”（样本中的单元数目）的增加，标准误 SE 也逐渐下降，尽管下降得相当缓慢。如果总体是相对同质的，标准误 SE 会很小；而总体的异质性程度通常可用标准差或某种类似的统计量，从样本数据中估计出来。整群样本特别是有很多群时，倾向于产生大的标准误 SE，尽管此类设计通常具有成本效率。

在实际工作中，非抽样误差【Non-Sampling Error】通常是更加严重的问题，但是同抽样误差相比，它更难以量化，受到的重视也更少。非抽样误差不可能通过增加样本量来控制。实际上，更大的样本量也更加难以控制。增加样本容量从抽样误差的角度看是有利的，但是从非抽样误差的角度看却可能适得其反。

非抽样误差本身可能分为三种，即选择偏倚【Selection Bias】、无回答偏倚【Nonresponse Bias】和回答偏倚【Response Bias】。我们依次讨论如下。

①选择偏倚是一种从样本中排除某种总体单元的系统趋向。在方便抽样中，选择偏差是一个主要的问题。如果是一个设计良好的概率抽样，就会有最小的选择偏差。这就是概率样本最主要的优点。

②一般而言，那些挂断你电话的人与那些愿意接受访谈的人是有区别的。这一区别即表明了无回答偏差。由于无回答偏倚，从回答人外推到无回答人是有问题的。如果回答率高（完成了大部分访谈），无回答偏差就会很小；如果回答率低，无回答偏倚就是一个需要考量的问题。

在本文写作期间，美国政府的调查——容许家庭户内的任意一位作为回答人——有超过 95%的应答率。美国最好的面对面的调查研究，即从家庭户中随机选取一名成年人进行访谈，可以有超过 80%的应答率。最好的电话调查【Telephone Survey】可以有接近 60%的应答率。而许多商业调查的应答率更低，足令人担心其推论的可靠性。

③回答人可能很容易由于访谈员的态度、问题的烦琐措辞，甚至两个问题的并列而不说真话。这是回答偏倚的常见来源。

就概率抽样而言，抽样误差有明确的界定。但是这个概念可以延展到方便样本吗？对此是有争议的（见下文）。概率样本成本很高，但是可以极小化选择偏倚，还可以为估计样本误差的可能影响提供基础。不过，回答偏倚和无回答偏倚同时会影响概率样本和方便样本。

用回答人替换无回答人

许多调查都有计划好的样本量，如果遇到了一位无回答人，即用一位回答人作替换。这可能有助于控制抽样误差，但是对于减少偏差却没有任何帮助。如果调查准备要从回答人外推到无回答人，就必须要知道遇到了多少位无回答人。

样本应当多大？

就这一常见问题，并没有一个确切的统计学上的答案。大样本有更小的抽样误差。另一方面，小样本更容易实施，会有更少的非抽样误差。大样本比小样本花费更多，一般而言，资源限制会决定样本量。如果做了试调查，就可能据此判断样本量大小对于最终的估计精度的意义。

如果研究的关注点在于相对误差，总体大小就很难成为一个决定因素。例如，对美国总统选举的民众选票的百分比统计分析——有 2 亿潜在的投票者——用一个有几千人的样本，就可以获得相当好的估计。当然，比起从在爱达荷州（Idaho）波伊西市

(Boise)的 20 万人中选取一个样本,从全美 2 亿人中选取一个样本需要付出更多的努力。

分层和权重

通常,还可以把样本框按“层”分割为不同的群组,再从每个层中各自抽出一个简单随机样本。如果层内的单元是相对同质的,这样做就会得到统计效率上的增益。其他种类的效率也会由此产生。如果我们对一个城市中的街区抽样,那么有些街区的人口可能是分散的。为了节约访谈者的时间,明智的做法是以一个比大密度人口集中街区低的比率选取此类街区作为样本。如果目标在于研究贫困决定因素,那么更好的做法是从位于更穷的邻里地区中抽出更多的街区作为样本。

如果用不同的比率从不同的层中抽取样本,分析过程也必须考虑抽样率。例如,Horvitz-Thompson 估计量就是以每个样本单元的选取概率的倒数作为每个单元的权重。这个估计量是无偏的,尽管它的方差可能很高。如果没能使用适当的权重,这通常就会引致偏差,而在某些情况下,偏差还会相当大(就方便样本而言,通过运用权重控制偏差是不可能的)。

针对同一目标总体,相较于以同量的简单随机样本为基础的估计量【Estimator】,基于某种复杂设计的估计量通常会有更大的方差:分群(层)是一个原因,权重的变异是另一个原因。复杂设计的方差同简单随机抽样的方差之比称为设计效应【Design Effects】。

比率和差值估计量

假设我们必须对大量的赔付申请进行稽核,进而确定它们总的过审价值,并将其同簿记价值(book value)相比较。对总的申请进行稽核会花费巨大,因此,我们只有取一个样本。同时,申请价值中相当大的百分比可能存在于一堆占比很小的申请之中。因此,我们可能需要给予那些大额申请更大的选取概率,给予小额申请更小的选取概率,然后运用权重进行调整。不过就目前而言,让我们只考虑一个简单随机样本。

假设,我们取样本申请中总的过审价值与总的簿记价值的比,再乘以所有申请的总的簿记价值。这就是一个对总申请中的总过审价值的“比率估计量”。比率估计量是有偏的,因为它们的分母是随机的,但是偏差的大小可以从数据中估计出来,并且常常可以通过减少抽样易变性获得抵消。比率估计量有广泛的应用。

较不常见的是差值估计量(difference estimator)。在上述赔付申请的例子中,我们可以就每一个样本的赔付申请,取其过审价值与簿记价值之差。再将样本平均数,即每一赔付申请钱数的均值,乘以所有赔付申请的总数目。这一估计量,即过审价值与簿记价值的总差值是无偏的,同比率估计量一样有用。

比率估计量及相似的估计量依赖于对取样的总体有额外的信息。在我们的例子中,要知道总体中赔付申请的数目,以及每个申请的簿记价值;而过审价值仅适用于样本。就分层而言,还需要总体的其他信息。可以用赔付申请的数目和赔付申请的簿记价值,根据申请额的大小分出几个不同的层。如果存在关于总体的相关额外信息,分层会增进估计精度。

计算标准误

就简单随机样本而言,样本的平均值是总体平均值的一个无偏估计值——假定回答偏倚和无回答偏差可以忽略不计。样

本平均值的标准误一般可通过用样本的标准差除以样本量的平方根得到很好的近似。

就复杂设计而言,不存在方差的简单公式;但是,可以用折刀法(jackknife)这类方法获得近似的方差(标准误是方差的平方根)。对于像比率估计量一类的非线性统计量,可以使用 delta 法。

抽样分布

此处我们仅考虑概率样本,并暂时搁置回答偏倚和无回答偏倚。一个估计量可以采用由不同样本(抽样变异)得到的不同的值;采用某个特定值的概率,在原则上可以由样本设计决定。该估计量的概率分布即它的抽样分布。该估计量的期望值即它的抽样分布的中心点,而该估计量的标准误即该分布的散布程度。从技术上讲,一个估计量中的偏差即其期望值与被估量的真值的差。

几个例子

1936 年,罗斯福进行第二次竞选,对手是艾尔弗·兰登。很多观察者期待罗斯福击败艾尔弗·兰登——但是《读者文摘》(*Literary Digest*)不作此期待,它预测罗斯福只能得到 43%的选票(选举中,罗斯福得到了 62%的选票)。文摘预测的基础是一个有 240 万人的巨大样本。其中,问题不在于抽样误差。问题来自无回答误差,而要找到它的来源,就需要考虑样本是如何被选出的。

文摘邮寄出了 1 000 万份调查问卷,得到了 240 万份的回复——为无回答偏差留下了相当大的余地。此外,据以邮寄问卷的邮寄地址清单根据汽车所有者和电话号码簿编辑而成,这就产生了另一偏差源。1936 年,汽车和电话还不像今天这样普及,所以文摘的邮寄地址清单中登录了过多的在大萧条时期能支付得起奢侈品的人。这就构成了选择偏差。

现在我们回到 1948 年,其时主要民意调查机构[包括盖洛普和罗珀]选出杜威做总统,而不是杜鲁门。其中一个庆祝获胜的新闻标题是:

统计学告诉罗珀,杜威当选!

样本很大——几万名回答人。问题仍是非抽样误差,即问题出在选取样本的方法上。它使用了配额抽样【Quota Sampling】。访谈员可以任意地选取他们喜欢的对象作为访谈对象,但是事先规定了某种数字配额。例如,一位访谈员必须选取 7 名男性和 6 名女性;男性中有 4 名必须是 40 岁以上的;等等。

配额的设定方式:在总计上,样本在性别、年龄和其他控制变量上非常接近于总体。但是问题在于谁会投票给杜威?在每一个样本类别中,总有些人比其他人更可能投票给共和党人。没有可能为可能的共和党投票人设置相应配额,因为在调查发生时期,他们的数量是未知的。而最终的结果是,访谈员更偏爱访谈共和党投票人而不是民主党投票人,这不仅在 1948 年是如此,在所有先前使用了同样方法的选举中,情况都是如此。

访谈员偏爱共和党投票人是选择偏倚的另一个例子。1936 年、1940 年和 1948 年,罗斯福赢得连续的胜利,民意调查中的选择偏差并不足以改变预测的方向。但 1948 年的选举是一场势均力敌的竞选,选择偏差使得民意调查的结果偏向了另一边。配额抽样看上去有道理,而且至今也被广泛地使用。不过 1948 年以来,概率抽样的优势应是众所周知了。

我们的最后一个例子是一项针对美国人口普查的校正方案。这是一个复杂的议题,但是简单地说,这个方案的内容是,在普查之后做一个特别的样本调查[事后确查调查(post enumeration survey)],以估计人口普查中的误差率。如果可以以充分的精度估出误差率,就能纠正它们。事后确查调查是一个分层街区样本,即依上文描述的方法选取的样本。样本量是巨大的(2000 年是 70 万人),同时抽样误差在合理的控制范围内。不过,相对人口普查中需要修补的小误差来说,非抽样误差仍然是个问题。对这一问题的各种讨论,可参见英伯的著作(Imber,2001)。

超总体模型

某些时候,人们分析方便样本的方式,就好像它们是从某个很大,但界定不甚明确的总体中选出的简单随机样本。这一不受支持的假定有时被称为超总体模型。过去人们常常作出这一假定,但是假定的频繁不能成为仍旧如此假定的合理依据,而其夸张的名称也不是如此假定的理由。假定总有其后果,因此,在深思后可能做出的假定是,默许方式不可能解决归纳问题。更多讨论,可参见伯克和费里德曼的文章(Berk & Freedman,1995)。

一个方便样本的标准误(SE)最好被视作一个没有意义的误差估计值:如果与事实相反,将其当作一个简单随机样本,其不确定性就会像 SE 一样被归之于随机性。然而,对 SE 的计算不应当将人们的注意力从非抽样误差中移开,因为后者仍然是抽样的首要考量(标准误测量的是抽样误差,一般不涉及偏差)。

一些实践上的建议

抽样调查研究并不容易,在下面的参考文献中可以找到一些有益的建议。调查者要在设计阶段下一番功夫。研究假设和目标总体都应当清楚地界定。如果对人实施访谈,访谈员需要得到培训和督导。调查过程的质量控制是根本。因此,相关文档和记录的保存也是关键。调查工具本身要有好的设计。问题应当简短、清晰;问题的措辞要能引出真实的答案,而不是迎合的回答。强烈建议调查前实施一次或多次试调查。有必要极小化无回答偏差;如果无回答是可察觉到的,就应选取一个无回答人的样本,并对之进行访谈。在最大的可能范围内,使用概率方法抽取样本。

——David A. Freedman

(王玥译　赵锋校)

参考文献

Berk, R. A., & Freedman, D. A. (1995). Statistical assumptions as empirical commitments. In T. G. Blomberg and S. Cohen (Eds.), *Law, punishment, and social control: Essays in honor of Sheldon Messinger* (pp. 245-258). New York: Aldine de Gruyter. 2nd ed. (2003) pp. 235-254.

Cochran, W. G. (1977). *Sampling techniques*, 3rd ed. New York: John Wiley & Sons.

Diamond, S. S. (2000). Reference guide on survey research. In *Reference manual on scientific evidence*, 2nd ed. (pp. 229-276). Washington, D. C.: Federal Judicial Center.

Freedman, D. A., Pisani, R., & Purves, R. A. (1998). *Statistics*, 3rd ed. New York: W. W. Norton, Inc.

Hansen, M. H., Hurwitz, W. N. & Madow, W. G. (1953). *Sample survey methods and theory*. New York: John Wiley & Sons.

Imber, J. B. (Ed.). (2001). Symposium: Population politics. *Society*, *39*, 3-53.

Kish, L. (1987). *Statistical design for research*. New York: John Wiley & Sons.

Sudman, S. (1976). *Applied sampling*. New York: Academic Press.

Zeisel, H. & Kaye, D. H. (1997). *Prove it with figures*. New York: Springer.

抽样偏倚(Sampling Bias)

参见抽样误差【Sampling Error】。

抽样分布(Sampling Distribution)

一个特定统计量的抽样分布是以一定抽样设计条件下的所有可能的样本为基础的该统计量的可能值的频次分布【Frequency Distribution】。这个分布很重要,因为若给定仅实施一次抽样调查,抽样分布即等同于一个概率分布【Probability Distribution】。因此,唯一实现的调查统计量就是得自于抽样分布的一个随机观测值。抽样分布的知识容许研究者去计算该统计量的值落在某一特定范围内的概率。

一个抽样分布的两个重要特征是它的位置和它的分散度【Dispersion】。如果分布【Distribution】的均值落在总体参数【Parameter】(样本统计量即是它的一个估计值)的真实值上,那么即说样本统计量是总体参数的一个无偏【Unbiased】估计量。

换言之,"平均"的样本估计值正好等于总体的参数值,但是它不必是每一次抽样的结果。分布的分散度越大,某个实现的统计量(即便它是一个无偏估计量)就越可能偏离总体参数到某个给定的量级。测量分散度的一种常见方式是计算分布的方差。样本设计者要努力极小化估计量的方差,因为这样就能极小化选中不适当样本——该样本会产生一个同总体参数有相当差异的估计值——的概率。总之,估计量的抽样分布在理想上应当是无偏的,还要有很小的方差。

抽样分布只有在使用了随机抽样【Random Sampling】的设计中才可以估出。如果是非概率设计,概率分布的概念就不适用。如果是随机抽样设计,(根据随机化推论)估计量就被视为是无偏的,所以就只需考量抽样分布的方差。

实践中,就众多调查统计量而言,假设样本量不是太小,抽样分布最终会趋近于正态分布【Normal Distribution】。结果是,可以利用正态分布的已知属性来构建估计值的大致的置信区间【Confidence Interval】。例如,95%的可能的估计值会落在分布均值的正负 1.96 个标准差的范围内(见图 1)。

因此,只需估计抽样分布的标准差【Standard Deviation】——更常见的说法是估计值的标准误【Standard Error】。然而,正态近似不一定普遍适用。特别是特定类型的估计值可能倾向于产生非对称的抽样分布。这种情况下,就必须使用其他方法来估计置信区间,如再抽样方法。

还需要认识到的一个要点是,在某个单独的抽样设计中,不同的样本统计量有不同的抽样分布。因此,即便在开展调查前就可能估出均值和方差的抽样分布,却不太可能选择一个对所有统计量都最理想的设计。抽样设计通常是一种妥协,其判

断标准在于被选中的抽样设计所生产的抽样分布对所需估计的大多数估计值是否可能足够好。

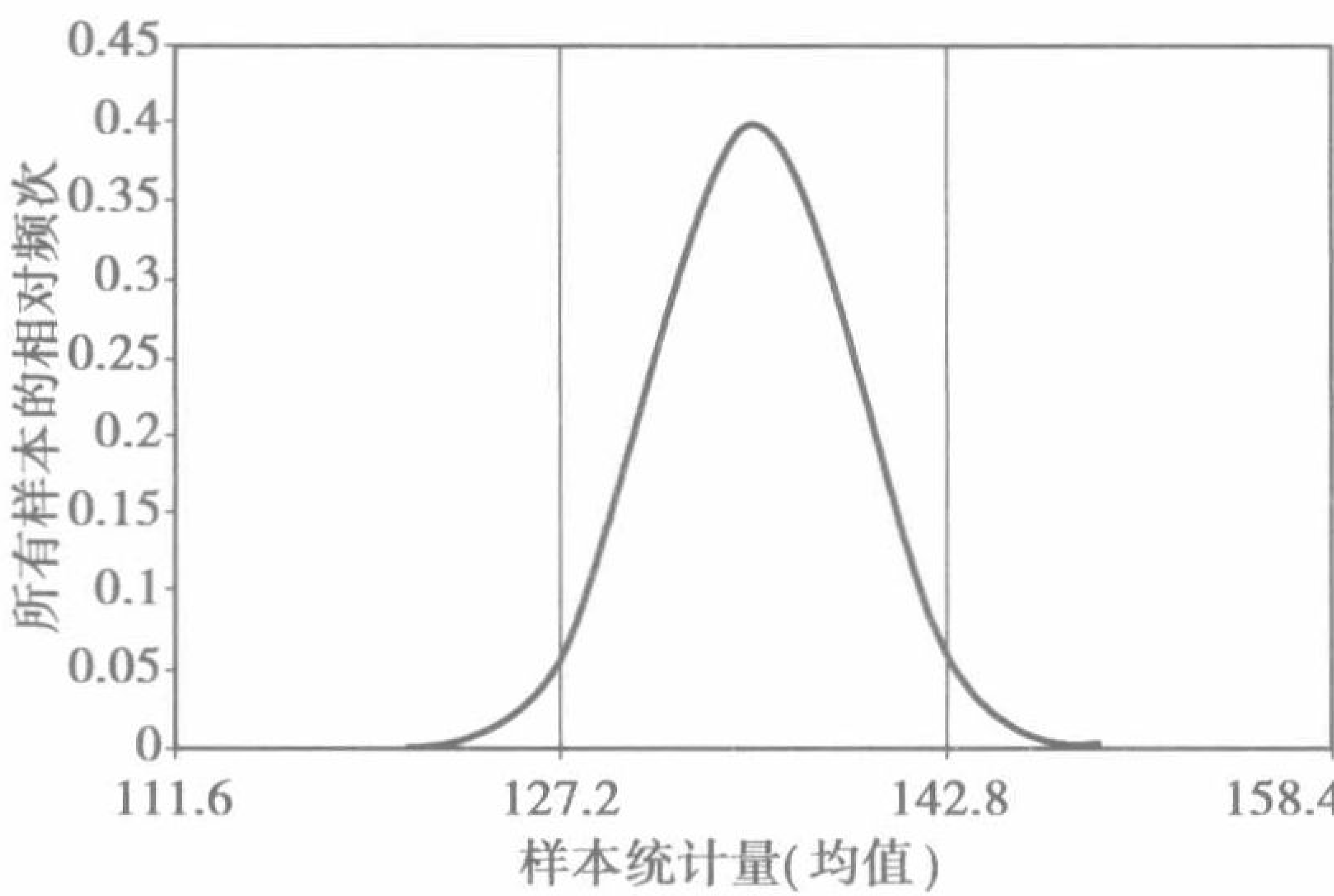


图 1 抽样分布

注:抽样分布的均值(期望值)是 135.0。两条垂直线分别表示均值的正负 1.96 个标准差(标准差为 3.98)。在可能获得的所有估计值中,2.5%的值小于 127.2,2.5%的值大于 142.8,剩下的 95%位于这两个值之间。

——Peter Lynn

(王玥译 赵锋校)

参考文献

Mohr, L. B. (1990). *Understanding significance testing*. Newbury Park, CA: Sage.

Moser, C. A., & Kalton, G. (1971). *Survey methods in social investigation* (2nd ed.). Aldershot, UK: Gower.

Stuart, A. (1984). *The ideas of sampling*. London: Griffin.

抽样误差(Sampling Error)

任何将抽样调查【Survey】作为研究某个较大兴趣总体【Population】的手段的尝试都将冒着选取的样本同总体特征不完全一致的风险。结果是,以样本为基础的估计值与它们所估计的总体参数【Parameters】所取得的数值并不一定相同。术语抽样误差即指这种样本估计值不同于真实的总体值的一般趋势。抽样设计的主要目标在于使得出现样本特征值非常不同于总体特征值的风险降到最小。

在使用随机抽样方法(参见随机抽样【Random Sampling】)进行定量样本调查的背景下,抽样误差还有更专门的意义。它是指根据某个特别设定的样本设计、某个特定的样本统计量同相应的总体参数间的期望的离差(expected deviation)。这一期望的离差由两部分组成,一部分可归之于随机的抽样误差(变异度,或方差【Variance】),还有一部分可归之于系统的抽样误差(偏倚【Bias】)。因此,对于无偏的样本设计(或严格地说,以样本为基础的无偏估计量)来说,抽样误差即等同于抽样方差。通常,把调查估计值(survey estimates)的标准误【Standard Error】(抽样方差的平方根)作为抽样误差的估计值的估计和表示。标准误可以用于构建置信区间【Confidence Interval】。更一般的情况下,均方误差【Mean Squared Error, MSE】被用来作为抽样误差的一个量度。均方误差定义为:

$$\begin{aligned} \mathrm{MSE}(y) &= E[y - E(y)]^2 + [E(y) - Y]^2 \\ &= \mathrm{Var}(y) + \mathrm{Bias}^2(y) \end{aligned} \quad (1)$$

其中,样本统计数 y 被用于估计总体参数

Y。MSE 是方差和估计值的偏差平方的合计。

当然，这个量仅适用于某种特定样本设计下某个特定的估计值。实践中，就大多数抽样估计值而言，由于偏倚的大小是未知的（如果它是已知的，就可以从抽样估计值中去除它，而无须使用一个偏倚的估计量），这个量也就是未知的。很大程度上由于这个原因，对抽样误差的估计通常就限于对某个估计值的方差的估计。同样，在许多情况下，可以合理地假定抽样偏倚为零，或可以忽略。假如关于调查数据的重要的设计特征，如整群、分层及设计权重等，不仅被指明，且已被适当地纳入计算的考量中，估计值的方差可以很容易地从调查数据中估得。对多数常用的样本设计而言，方差的形式是已知的；因此，可以用标准化的估计方法。对于那些特别复杂的设计，或缺乏清楚说明的设计，就有必要使用某种再抽样【Replication】方法，对样本统计量的方差予以经验估计。这类方法，如刀切法【Jackknife Method】、自助抽样推论法【Bootstrapping】以及平衡的多次再抽样，正越来越多地出现在统计软件包中。

抽样误差只是总调查误差的一个组成部分。一个统计量的总调查误差是该统计量与相应的总体参数间的期望的偏差。它不仅混有抽样带来的误差，还包括可归之于无覆盖、无回答【Nonresponse】带来的误差；以及测量（访谈员误差、回答人误差、测量工具误差、操作误差）的误差。在许多情况下，尽管抽样偏倚可能是可以忽略不计的，但是其他来源的偏倚可能有很大影响。由此，将标准误作为抽样估计值的精度的指针可能是有误导性的。

——Peter Lynn

（王玥译　赵锋校）

参考文献

Scheaffer, R. L., Mendenhall, W., & Ott, L. (1990). *Elementary survey sampling*. Boston: PWS-Kent.

Stuart, A. (1984). *The ideas of sampling*. London: Griffin.

抽样分数(Sampling Fraction)

抽样分数是指被选出的构成调查样本的单位占抽样框【Sampling Frame】的分数或比例。如果抽样框中有 N 个单位，n 个在样本中，那么抽样分数就是 n/N。

抽样分数有时会与"选中概率(selection probability)"混淆。抽样分数为 n/N 并不一定意味着每个单位的选中概率是 n/N。可能 N 中不同单位有不同的选中概率。抽样分数仅指最终组成样本的全体单位所占的整体分数，并不涉及它们被选取的方式。

一些调查对总体【Population】的不同部分(层)使用不同的抽样分数。这就是"不定比分层抽样(disproportionate stratified sampling)"。它常被用于增加某些群体的样本【Sample】量，而这些群体是总体中的少数，且具有特别的分析价值。给予具有较大总体方差的层以更大的抽样分数还能增加估计值的精度(参见分层样本【Stratified Sample】)。

在系统抽样的情况下，抽样分数既反映出抽样过程的结果，也反映出过程本身，同时，它也等同于选中概率。抽样分数 $1/A$ 意味着抽样清单上的每一个第 A 个单位是被(或将被)选中的。A 为抽样分数的倒数，被

称作抽样间距（参见系统抽样【Systematic Sampling】）。

除非它非常大（例如，超过 0.05），抽样分数对于抽样估计值的精度的作用是可忽略的。令人惊讶的结果是：例如，无论是对一个有 10 亿人口的总体，还是对 100 万人口的总体的特征做估计，1 000 个样本有同样好的效果。

——Peter Lynn

（王玥译　赵锋校）

参考文献

Barnett, V. (1974). *Elements of sampling theory*. Sevenoaks, UK: Hodder & Stoughton.

Cochran, W. G. (1973). *Sampling techniques* (3rd ed.). New York: Wiley.

Moser, C. A., & Kalton, G. (1971). *Survey methods in social investigation* (2nd ed.). Aldershot, UK: Gower.

抽样框（Sampling Frame）

抽样框是一个用以确定研究者的目标总体【Population】的清单或其他的类似设计。抽样框确定了一个被抽取单元的集合，从中研究者可以选出目标总体的样本【Sample】。在社会科学研究中，因为一名研究者很少会直接接触到全部的兴趣总体，所以他必须依赖于一个抽样框来代表兴趣总体的全部单元。

一般而言，抽样框可分为两种：清单式的和非清单式的。清单式抽样框的例子包括一个城镇的注册选民的清单、列于某一地方电话号码簿上的居民，或注册一个课程的学生的花名册。非清单式抽样框的例子包括用于随机数字拨号【Random Digit Dialing, RDD】的一组电话号码，以及用于地区概率抽样【Probability Sampling】的某个确定范围内的所有住宅单元。随机数字拨号的抽样框是专门为电话调查【Telephone Survey】研究——访谈员会致电给电话号码被选中的家庭，并进行访谈——而设计的。地区概率抽样框主要用于访谈员在场的抽样调查【Survey】，即访谈员登门拜访每个被抽中的住宅单元，进而完成访谈。大部分清单式样本可以用于任一一种数据收集模式，包括访谈员在场调查、电话调查和邮寄调查。

一个好的抽样框最重要的两个目标是全面（comprehensiveness）和准确。全面是指抽样框覆盖整个目标总体的程度。例如，地方电话号码簿上的居民清单一般不是一个能代表这一地区全体居民的适当的抽样框。电话号码簿并没有包括那些电话号码未被列入的居民，或那些近期才开始使用电话服务的居民。相反，RDD 抽样框可以提供一组当下的、同某一给定电话交换机相连的、所有列入或非列入的电话号码。尽管 RDD 抽样框在这一点上要优于电话号码簿，但是这两个抽样框都不包括那些没有使用电话服务的居民。

准确是指抽样框包括其所覆盖目标总体单元的正确信息的程度。由于数据记录的错误、过期信息以及其他问题，电话号码簿、城镇清单和其他种类的清单经常包括那些不准确的信息。即便某个抽样框可能是全面的，这类清单所提供的信息的准确度也会影响该抽样框的质量。

缺少十分全面和准确的抽样框可能是任何一个研究项目中偏倚的一个主要来源。当抽样框排除了那些属于目标总体的单元或包括了那些不属于目标总体的单元时，偏倚就会产生。当不正确地被排除或被包括

的单元同目标总体的区别达到一定程度时,这种差别就会产生偏倚。如果研究者可以识别那些被错误地包括在抽样框中的单元,那么研究者就可以简单地将它们从抽样过程剔除。然而,关于那些被排除于样本框单元的目标总体成员,研究者通常仅有很少的信息。那些不能被识别且再次被覆盖的丢失的样本框单元就并没有被选入样本的机会。这样的一个后果是,从带有丢失单元的抽样框中抽取的样本通常不能充分代表目标总体。

——Douglas B. Currivan

(王玥译 赵锋校)

参考文献

Fowler, F. J., Jr. (2002). *Survey research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Kalton, G. (1983). *Introduction to survey sampling*. Beverly Hills, CA: Sage.

定性研究的抽样(Sampling in Qualitative Research)

抽样【Sampling】是根据参与者的特征,以满足研究的理论需要的方式,有意选择最适合的参与者到研究中。在定性质询中,注意样本的选取不仅对研究的严谨有重要意义,还须要在研究的全程一以贯之。

样本的类型

样本可以根据从总体中选取参与者的方式予以分类。在定性调查中,选取的过程是有意的(deliberate),而不是随机的。

方便样本

方便抽样是通过纳入那些易于接触且符合研究标准的参与者而选取的样本。在调查开始时,当调查者不知道样本选择的适当的特征标准时,可使用方便样本【Convenience Sample】,或者当能接触到的参与者数量很少时,也可以使用它。例如,想要系统地研究一下曾做过心脏手术的儿童的经历,或是依赖呼吸器的中小学生,这类人在既定地理区域这一病患总体数目很低。如果所有可接触的参与者都已被纳入研究(例如,在一间特定教室中的所有儿童),这就是完全样本。

选取某人作为参与者,将其纳入一个定性访谈时,最低限度要求,参与者必须愿意且能够对经验进行反思,明白地表达他们的经验,还要有充分的时间,在访谈时不被打扰。若是完全样本,可能出现的问题是:如果一名参与者受到邀请,或志愿参与访谈时,研究者不知道参与者是否符合"好的"参与者的最低标准。如果将所有的访谈参与者都纳入初始样本,我们就没使用任何筛选程序。另一方面,如果研究者开始进行访谈,却发现一个参与者不是一个好的知情人,既没有预期的经验,对正在寻求的信息也一无所知,我们就不得不进行二次选取(Morse, 1989)。在这种情况下,研究者要礼貌地完成访谈。事后,将录音带做好标记,放在一旁。研究者不必浪费时间和研究经费转录【Transcription】这些录音带。既不要将访谈资料并入数据集,也无须删除录音。如果录音带还有意义,以后可能用得上,就将它保存下来。

理论样本

这是根据研究的理论需要选取参与者。如果研究者已经确认一个需要比较的研究问题,并知道所需展开比较的维度,就可以在资料收集的开始阶段使用理论抽样

【Theoretical Sampling】。举个例子,研究者可能用一个两组设计来比较参与者,比如,比较连续领取失业救济的单亲母亲和那些阶段性工作偶尔申请救济的单亲母亲。理论抽样可用于两组资料的比较及两组人之间差异的辨识。

在分析过程中,类别和主题逐渐形成,而某些分析领域可能比其他充实得慢些,或显得较为单薄。一份理论样本是有意寻求的参与者,即那些具有特定知识或特征,而他们的知识和特征将有助于完善带有不充分资料的类别,也就是所谓的分析的稀薄领域,为其补充资料。另一种情况是调查者想要检验某一推测或假设【Hypothesis】,就可以寻找和访谈那些特征可以与其他参与者比较或对照的参与者。一个例子可能是,单亲失业母亲的先前研究发现,那些反复工作的母亲比那些一直失业的母亲有更强的适应力。因为这个样本的参与者太少不足以证明突现的研究发现,所以研究者必须有意地寻找符合标准的参与者,以充实这些类别。这一类型的理论抽样,即根据个人特征而有意地寻找参与者,有时也被称为目的抽样【Purposive Sampling】。

负面个案

负面个案的抽样是有意寻找那些对正在发展的理论框架而言属于例外的参与者。如果我们继续以失业单亲母亲为例子,一个可能出现的负面个案是,一个母亲,阶段性地返回岗位,但是没有表现出适应力的明显特征。为了正在发展的理论的完整和有效,必须有目地选取负面个案,直到否定类别变得充实,而且这些例外的逻辑解释也被纳入理论。

定性抽样的进程

如前所述,在定性质询中,抽样持续于整个研究过程,从而在资料收集和分析之间形成了一条反应链。起初,调查者可能实行开放抽样(open sampling)——在有限的主题内,有意寻找尽可能大的变异度。这种变异度使得问题或现象的范围被勾勒出。以后,随着抽样的进行、理解的获得,以及类别的浮现,抽样策略可能改作理论抽样,以期证实突现的理论陈述。于是,抽样一直持续直至理论变得深厚、稠密、饱和。

至此,我们都是把抽样当作选取研究参与者的过程。在定性质询中,抽样也出现在分析过程中对特定资料的选取和注意上。在定性分析中,无须同样地对待所有的资料——研究者需留意某些资料,因为这些资料可能更加丰富,或可以提供更多有助于发展理论洞见的解释。其他资料可能被用于支持或反驳这些正在发展的概念,也有些其他资料可能不包含任何相关的信息,因此在分析的某一特定阶段上可以不必留意,甚至将其忽略。在写作阶段,研究者则有意选择最佳引语作为例子,以此说明正在论证的观点。

为什么随机样本不适合

随机性意味着一个确定群体中的所有人都是研究的潜在参与者。然而,定性研究者知道并非所有的参与者都是好的参与者。我们知道有些参与者可能同他人相比不甚愿意参与研究(他们可能没有时间,可能不愿意考虑相关主题,可能生病了,诸如此类),还有些人同他人比,可能表述更清楚,更善于反思。最重要的是,不是所有的参与者都有同样的生活经历能引起访谈者的兴趣。因为定性资料的获得代价高昂(访谈非常费时),还必须转录;文本分析起来既不方便也耗费时间,所以定性研究者不可能冒险去收集同主题关联不密或脱离主题的资料。

另外,随机化过程也不能给予研究者机会,以控制他们欲收集的有用资料的总量。因为关于某一主题的知识在人群中不是随

机分布的，所以根据机会来收集资料可能导致在通常已知的主题上收集了过多的资料，而在少有涉及的主题上资料则不够充分。于是，在某些领域拥有的资料过多，另一些领域则不足。因此，人们常发现，大量资料在这些主题上频繁地出现，可能压制和弱化某些稀见但同样重要的特征，从而在实际上使分析出现偏差和失去效度。

样本的评估

定性研究的新手会有如下问题：需要多少参与者？需要多少观察个案？需要多少组资料？资料的数量依赖许多的因素，包括资料的质量、现象的性质及研究的范围。认真考虑以下原则：

1.研究的范围越广，所需资料的数量越大。

调查新手的一个常见的困境是，研究开始前，没有界定问题，以致他或她试图研究所有的事情。当然，这就意味着需要更多的参与者和更多的资料才能达到饱和。另一方面，草率地规定主题可能导致排除了问题的重要的方面或简单化了问题，同样是无效的。

2.访谈的质量越高，需要的访谈就越少。

理想情况下，应当就一个合适的样本实施访谈。如果样本是仔细选取的，同时，如果所获信息是极少与主旨无关的，又能启发洞见，且富于细节，那么这名调查者就可以相较于另一种情况更快地结束抽样——另一种情况是，他或她得到的访谈意思含混、不切主题、答案模棱两可或矛盾。

3.在每名参与者提供的资料数量同参与者的数量间存在着反向关系。

相对于使用半结构式访谈【Semistructured Interview】，即只去问参与者一定数量的主干问题，深度访谈法之类的方法，如对同一访谈参与者进行3次、每次1小时以上的互动访谈，用很少的参与者，就可以很快实现饱和。

如何分辨抽样已经足够了？

一句箴言说，当定性研究者感觉厌倦时，他或她就会停止抽样（Morse & Richards, 2002）。或者，若没有新的资料出现，或研究者已经“听过这一切”了，饱和就实现了。如果样本充分且适当，研究的主题又有限，饱和就会更快地出现。

充分

充分指足够多的资料。在资料集中，资料必须重复出现。此即，只在某人那里听过一次的东西是不能令人满意的，这些想法应当得到他人的印证，或通过其他途径（档案资料、观察资料等）予以印证。这些印证可能形式相同或形式不同，比如，不同的故事，它们例示了研究者想要制造出的相似点。充分的资料对于概念和理论的构建是必需的。实际上，用单薄的、不充分的资料去发展类别比用充分的资料更加困难，因为，在没有重复【Replication】的情况下，看出模式是困难的，同时，如果资料存在缺口，也难以看出其中的主旨。

适当

适当指有意选取最佳参与者参与研究。在定性质询中，适当的抽样（使用上文的目的抽样和理论抽样）持续于整个研究过程。

——Janice M. Morse

（王玥译　赵锋校）

参考文献

Morse, J. M. (1989). Strategies for sampling. In J. Morse (Ed.), *Qualitative nursing research: A contemporary dialogue* (pp. 117-131). Rockville, MD: Aspen.

Morse, J. M., & Richards, L. (2002). *Read me first*

for a qualitative guide to qualitative methods. Thousand Oaks, CA: Sage.

抽样变异(Sampling Variability)

参见抽样误差【Sampling Error】。

放回(无放回)抽样(Sampling With [or Without] Replacement)

简单随机样本【Simple Random Sampling】按照惯例是指无放回抽取的样本。无放回抽样【Sampling】意指,从总体选择一个单位放进样本时,这个单位不会再放回到样本池。在无放回抽样过程中,总体的每个单位被选作样本的机会有且仅有一次。例如,如果我们要从 999 个球中抽取 10 个球做样本,我们可以把所有的球放入一个箱子,同时在每个球上都标上 1~999 的一个唯一的数码。然后,摇晃箱子,以公平的程序抽取第一个球。如果我们要抽一个 10 个球的样本,第一个球的数码——此例中,即 071——被记下来,然后把这个球放在一边。使用同样的程序抽取其他 9 个球,而且每次都把选中的球放在一边。因为被选作样本的球都被放在一边,这就是一个无放回抽样。传统上,多数以社会科学研究为目的的样本都假定是无放回抽样。

这一程序尽管看起来符合常识,即其作用看上去在于勿使所研究总体中的某个成员在一个单独的样本中出现不止一次,然而另一种替代的程序,放回抽样则具有一些非常好的性质,而且对统计理论来说,实际上非常重要。像先前的例子一样,同样使用摇晃球的箱子放回抽样,首先选出一个球——这次,比如说,第一个是 802——记下这个号码,再将球放回箱子。如此一来,802 就有再次被选取的资格。从统计理论的观点出发,这种程序的一个好处在于,为选取每一个样本成员的每次抽取相对于所有其他总体单位的选取都是完全独立的。在每次抽取中,总体的每个成员都可能被选中,因此,对每次抽取而言,选中的概率【Probability】相同。

在无放回抽样中,第一次抽取后的每一次抽取都依赖于先前的抽取。在选取无放回概率样本时,尽管未来选取对先前抽取的依赖度相对较小,有些问题仍应予以考虑。例如,在随机的实验【Experiment】中选取一个控制组【Control Group】时,这样做就会产生独立性的问题。如果先选取处理【Treatment】组,那么,从技术角度来讲,控制组样本的选取就依赖于处理组的选取。如果随机化程序并不是完全没有偏倚的,好像 1969 年美国军队乐透开奖出现的状况一样(McKean, 1987, p. 75),这就可能产生偏倚。在大多数情况下,通过检验处理组和控制组在某些特征上的差异——根据一定的理论,这些特征与结果变量有关联——可以核查此类偏倚是否出现。这些差异可以说明,偏倚有可能影响实验结果,但是不能确定地被否证。

另一个具有实践含义的考量是计算无放回样本的标准误【Standard Error】。从技术上讲,当抽取一个无放回样本时,为计算样本的标准误,需要一个有限总体限制(finite population control, FPC)。简单随机样本的标准误公式如下:

$$s_{(e)} = (1 - f)^{\frac{1}{2}} s / n^{\frac{1}{2}}$$

其中：

$s_{(e)}$——估计量 e 的标准误；

$1-f$——有限总体纠正因子；

f——抽样分数，或 n/N；

s——标准离差的估计值；

n——样本量。

有限总体限制的影响依赖于总体的规模和样本相对于总体的规模。一般而言，当样本占总体比例少于 5%时，FPC 没有实际重要性（Henry, 1990; Kish, 1965）。例如，假如从一个学区中的 1 万名学生总体中抽取 1 200 个样本，FPC 会减少抽样误差到大约0.94，这是一个相当大的量。然而，假如同样数量的样本是从某个州的 140 万名学生中抽取，FPC 会减少抽样误差到0.999 6，而这是一个极小的量，实际上，可以认为FPC = 1。社会科学研究的多数情形中，FPC是可以忽略的，但是，从某个小总体抽取一个样本时，它可以显著地减少无放回样本的抽样误差。

——Gary T. Henry

（王玥译　赵锋校）

参考文献

Henry, G. T. (1990). *Practical sampling*. Newbury Park, CA: Sage.

Kish, L. (1965). *Survey sampling*. New York: Wiley.

McKean, K. (1987, January). The orderly pursuit of pure disorder. *Discover*, pp. 72-81.

SAS

SAS（统计分析系统）是一个广泛用于各种学科（包括社会科学）的通用软件包。更多信息可参见软件相关的网站。

——Tim Futing Liao

（马妍译　赵锋校）

饱和模型（Saturated Model）

饱和模型与数据值有同样多的参数。饱和模型的预测值或拟合值等于数据值；此即，饱和模型可以完全再制数据。单个观测值或单个数据点可被视为一条信息；因此，在一个有 n 个数据点的样本中，存在着 n 条信息。根据 n 个数据点算出的每一个统计量或每一参数都会用去一条信息。当一个模型的参数数量与数据中的信息条数相等时，该模型即是饱和的，而且可以说这个模型有 0 个自由度【Degrees of Freedom】。

举个例子，设想一下我们从某个总体中得到了一个含有 3 个观测值的样本：$x_1 = 3$，$x_2 = 5$，$x_3 = 10$，然后，想用模型 $x_i = \mu + \beta_i$ 拟合数据，其中，μ 和 β_i 是模型的参数。β_i 对每个观测值可能有不同的取值（如 $i = 1, 2, 3$），而 μ 的一个合理估计是样本的均值；亦即，$\bar{x} = (3+5+10)/3 = 6$，这里用掉了一条信息，还留下了两条信息来估计 β_i。利用剩余的两条信息，我们可以通过取一个数据值与样本均值的差来估计两个 β_i。比如，$\hat{\beta}_1 = x_1 - \bar{x} = 3 - 6 = -3$ 和 $\hat{\beta}_2 = x_2 - \bar{x} = 5 - 6 = -1$。现在我们有 3 个估计参数：$\bar{x}$，$\hat{\beta}_1$ 和 $\hat{\beta}_2$，它们消耗了

数据的所有信息。最后一个参数 β_3 的值运用$\hat{\beta}_1$和$\hat{\beta}_2$来计算，根据数据值和它们均值的差的和为0来计算。此即

$$\sum_{i=1}^{3}(x_i-\bar{x})=(x_1-\bar{x})+(x_2-\bar{x})+(x_3-\bar{x})$$
$$=\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_3=0$$

解这个等式得到$\hat{\beta}_3$，$\hat{\beta}_3=-(\hat{\beta}_1+\hat{\beta}_2)=-(-3-1)=4$。利用模型估得的参数，可以得到同数据值完全相等的预测值，如下：

$$\hat{x}_1=6-3=3$$
$$\hat{x}_2=6-1=5$$
$$\hat{x}_3=6+4=10$$

模型可用以拟合数据，从而给出数据的综括统计或数据描述，也可以生成一个能预测未来值的方程，还可以用以检验关于总体（数据即从中抽取）的假设。一般而言，人们偏爱更加简单的模型，因为它们平滑掉了数据中存在的非系统的变化，而这些非系统的变化则来源于数据仅是一个总体的一个样本。虽然饱和模型极其精确，但是它们无法区分数据中的非系统的变化和系统的变化。

对非饱和模型的整体拟合优度【Goodness-of-Fit】的统计检验，饱和模型发挥着重要作用。所有的非饱和模型都是饱和模型的特例。考虑检验模型对数据的拟合优度的两个常用统计量：皮尔逊卡方统计量【Pearson's Chi-Square Statistic】

$$\chi^2=\sum_{i=1}^{N}\frac{(x_i-\hat{x}_i)^2}{\hat{x}_i}$$

和似然比统计量（likelihood ratio statistic）

$$L^2=\sum_{i=1}^{N}x_i\log\left(\frac{x_i}{\hat{x}_i}\right)$$

其中，x_i 等于个案 i 的一个观测值；$\hat{x}_i$ 等于从一个非饱和模型计算的预测值或者模型的测量值，而 N 等于观测值总量。在两个例子中，非饱和模型的预测值都被用以同数据值相比较，而后者是饱和模型的预测值。

——Carolyn J. Anderson

（马妍译　赵锋校）

参考文献

Agresti, A. (1996). *An introduction to categorical data analysis.* New York: Wiley.

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley.

Powers, D. A., & Xie, Y. (2000). *Statistical methods for categorical data.* San Diego, CA: Academic Press.

量表（Scale）

一份量表是对某些潜在概念【Concept】的一个复合量尺。社会科学中的许多现象涉及无法用一个单一变量【Variable】直接测量的理论构念【Construct】。为了理解这些现象的原因、效应和含义，研究者必须发展出这些构念的一组有效且可信的经验指标。这一经验指标即称作量表。在这个意义上，量表由一组能经验地捕捉到理论构念的本质意义的可测题项构成。

好的量表是数据简化的工具，即简化了数据中的信息，并用一个复合的量衡量一个构念的方向和强度。量表通常以定序

【Ordinal】测量层次来构建,当然也有部分量表可能是定距【Interval】层次的。因为量表由某个特定现象的多个指标构成,所以它们会同时增加信度和效度,而这是将单个变量用于分析所难以达到的。

有几种可能的模型可用于将诸题项合并到一张量表中。这些模型包括瑟斯通【Thurstone】量表、李克特【Likert】量表和哥特曼【Guttman】量表。方法的选择取决于量表想要达到的目的。

度量【Scaling】有3项相关但又有所区别的目的。首先,度量的意图可能在于检验某个特定假设【Hypothesis】(如假设一个单一的维度,政党身份认同会影响选举人对总统候选人的选择)。在这种情况下,度量模型被用作一个标准,以评估一个给定的观测数据集对于一个特定的模型的相对拟合。

度量的另一个目的是描述一个数据的结构。如果一个政治学家试图发现美国选举制度的潜在维度(如社会的、经济的、文化的等),这就是一个解释性研究而非假设检验法。

度量的第三种目的是构建一个量尺,从而使得个体能够被置于这把量尺之上,并能将他们在该量表上的得分与其他感兴趣的测量联系在一起。

度量模型可能用来测量人群、刺激或者同时测量两者。李克特量表是一种只测量受试者的度量模型。广义地说,任何通过累加诸题项的回答得分的量表均被称为李克特量表或者"累加"量表,或者一个线性复合量。数学课的一次测验就是一个线性复合量的示例。量表(测验)得分是通过累加对个别题项(问题)正确回答的数量得到的。与任何单一题目(问题)相比,量表(测试)中的复合得分是一个更好的学生对知识掌握程度的指标。

当各种刺激中没有明显的逻辑结构时,瑟斯通的兴趣在于衡量和比较诸刺激。瑟斯通量表模型试图识别诸刺激间的一个经验结构。为了达到该目的,瑟斯通量表运用人们的判断。受试者个体会被给予一组陈述,可能多达100个。这些"法官"根据一个潜在的维度对陈述进行排序。那些在"法官"中得到最大多数一致同意的陈述被选中作为量表的题项。这些保留下的题项(如20个)应当涵盖整个潜在的连续统。接着,这些题项会呈给受试者,并请受试者找出那些他们接受或同意的陈述。这些被选中的题项的平均量表分值代表了个人对问题中对象的态度。

最后要讨论的量表类型是哥特曼量表(或者称为由易到难分析或渐增度量)。它是根据某些潜在的渐增维度同时对题项和受试者进行排序的一种程序。其发展目的在于就一组题项的单维性给出一个经验检验。这一方法的潜在逻辑是渐增的逻辑;此即,知道了一位受试者在量表上的得分,就知道了受试者对每个题项的回答。如果受试者的得分是3,那么研究者不仅知道受试者正确地回答了3道题,而且还知道这3道题是第一题、第二题和第三题。多数哥特曼量表分析在于考虑是什么导致了从完美量尺的偏离。

这里仅是对量表的粗略讨论,包括它们是什么以及用它们来达到什么目的。量表在社会研究中有广泛应用,同时对它们的构建和解释也有许多研究。

——Edward G. Carmines
James Woods
(马妍译　赵锋校)

参考文献

Carmines, E. G., & McIver, J. P. (1981). *Unidimensional scaling* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-024). Beverly Hills, CA: Sage.

Guttman, L. L. (1944). A basis for scaling qualitative data. *American Sociological Review*, 9, 139-150.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *140*, 44-53.
Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*, 273-286.
Thurstone, L. L. (1929). Fechner's law and the method of equalappearing intervals. *Journal of Experimental Psychology*, *12*, 214-224.
Thurstone, L. L., & Chave, E. J. (1929). *The measurement of attitudes*. Chicago: University of Chicago Press.

度量(Scaling)

度量方法由给物的属性或特征赋以数值的过程组成。就这一点而言,"度量"似乎非常相似于"测量"(测量通常也界定为"给物赋以数值")。两者之间的差异是微妙的,但仍是重要的:测量常可简化为某类物理计数操作。例如,将一定数量统一长度的杆子(或者在一根木棒上有间隔一致的刻度)置于一个物体边,跨过物的一端到另一端的整个距离,从而根据杆的数量确定该物的宽度。同样,可以将一个物体放在天平的一侧,并根据平衡天平另一侧所需统一的砝码的数量,确定该物的质量。

一般目的

与此相反,度量用于量化那些无法直接计数的现象。例如,度量技术被用于测量调查回答人的态度、犯罪的严重性、职业声望以及立法者的意识形态。在上述每一种情况下,我们相信问题中的现象(如态度、严重性、声望、意识形态)确实存在。但是,这些东西中没有一个可以像估算一把尺子或一架天平上的刻度那样"被数出"。度量方法的存在就是为了这类现象。

没有一种单一的技术被称为"度量"。不过存在着各种各样的度量模型。这些模型的共同特征是,它们都试图通过把输入数据矩阵(the input data matrix)中的元素(即矩阵中的观测、刺激、变量等)描绘成一个几何空间(geometric space)中的对象(一般是点或向量),来复现真实世界的观测中存在的系统结构(systematic structure)。因此,一个量尺就是数据的一个抽象模型,而数据自身则通过经验观测收集而来。这一路径,即研究者从观测到数据,再到几何模型的逐步进展要依赖于具体的度量策略。

概言之,实施一个度量分析类似于做一个拼图游戏。像拼图中的小块,单个的数据点仅提供了零碎和有限的信息。当我们把一个个图块拼到一起时,结果是一幅观测者可以领悟和理解的图。同样,当观测值被合成某种几何结构时,作为结果的"图"会显出物在数据矩阵中的相对位置,从而有助于分析者理解数据本身的内容和含义。

数据和度量方法

针对一种特定的数据分析条件,人们应当如何选择一个具体的度量模型呢?答案取决于研究者对输入数据矩阵中包含的信息的解释,以及他或她想要构建并以此呈现这些信息的模型类型。两种最常见的数据矩阵类型是多元(multivariate)和相似(similarities)。多元数据是指一个数据矩阵其上的行和列分别代表不同的事物(通常,行代表第几个观测对象,列代表变量),而矩阵格的输入给出了两个事物交汇的一个集合的得分(如某一观测对象在某个变量上的

取值)。例如,调查回答人对现任总统10个特征的逐一评定即是一个多元数据的例子(数据矩阵上的行代表回答人——观测对象,列代表总统的某一特征——变量)。

就相似数据而言,矩阵的行和列代表同一事物(因此,矩阵实际上是正方形的),同时,矩阵格的输入表示行输入和列输入同彼此相似的程度。例如,一个实验受试者对10种食物的口味的相似程度的看法。但是相似数据,除了事物间实际的相似度以外,还可以呈现许多其他的东西。从度量的目的出发,诸多现象,如距离、协方差以及不同对象间享有共同特征的程度等,都可解释为相似数据。也许我们最熟悉的相似数据类型是相关【Correlation】系数矩阵(它测度诸变量间的两两"相似度")。

如果研究者拥有多元数据并且想获得关于行或列的一个单维尺度,那么李克特量表【Likert Scaling】(也称为累加评级量表【Summated Rating Scaling】)会比较适合。如果目的是行和列的一个单维呈现,那么哥特曼度量【Guttman Scaling】将会是一种有效的技术。如果分析者的兴趣在于在多维空间上定位行和列的输入,那么可以使用主成分分析【Principal Components Analysis】。再看相似数据,如果输入是一个相关系数矩阵【Matrix】,那么因子分析【Factor Analysis】通常是适宜的模型:相关系数被模型为向量(向量代表变量)之间角度的函数。多维度量【Multidimensional Scaling, MDS】常用于绝大多数其他类型的相似数据:在这种情况下,相似度被模型为某个几何空间上的点间距离(较小的距离对应较大的相似度)。

虽然不太常见,但除了多元和相似数据之外还有多种其他类型的数据——还有许多适于它们的其他度量方法。例如,有一个数据矩阵,它的行和列代表同一事物,而矩阵格的输入则给出,在某一具体特征上,行事物超过列事物的程度(或者相反),这种数据有时被称为刺激比较(stimulus comparison)数据。这类信息可以使用几种瑟斯通度量【Thurstone Scaling】技术中的任一种来度量。一个矩阵的行和列代表不同的事物,而矩阵格的输入表示每个行元素对每个列元素的接近度,这种数据有时被称为优先选择(preferential choice)数据。这种矩阵也可用作进一步分析的数据输入。

数据精炼

基于四个互补的理由,研究者倾向于使用度量过程。第一,为了数据精炼。分析者经常面对大量的信息,而且信息经常多得不易理解。因此,度量方法可用于"萃取"信息,并以一种更有效的形式表现出它最基本的构成。这样做时,最终的量尺常会生成一个综括量,而这个量要优于组成这个量的各成分变量中的任意一个。所以会出现这个结果是因为量表中某个题项的强度可以弥补其他题项的不足。因此,测量误差的效应被极小化了。

例如,一项民意测验可能就回答人对现任总统的态度,问他们几个问题。然后,可以将这些问题的回答组合成一个对总统态度的单一量尺。这么做的动机是在于:(1)单一变量比多个变量更容易进行分析;以及(2)相比较于任意一个单独的问题,这一量尺或许能给出回答人对总统的所有感受的一个更准确的呈现。

维数

第二,度量方法也可用于估计数据的维数。从实质的角度看,"维数"是指存在于所考察的事物之中,诸种分离且令人感兴趣

的变异源的数目。从分析的角度看,维数是指为了给出一个几何元素(还是点或向量)的精确图示——用以再现所考察的事物,所需要的坐标轴的数目。此外,所有的度量方法都试图用一个几何空间再现事物。空间的维对应潜在事物之下的变异性来源。

有时,事物间的变异性来源是事先已知的。因此,研究者可以在确定事物的点的位置之前,设定空间的维数。不仅就物理测量而言,情况如此,就所有其他的情况——在预先确定的标准的基础上,将事物予以对比,相对而言,情况也同样如此。例如,一项算术测验可被视为试图沿着某个单一维度,确定一组事物(学生)的位置的一个度量分析。

在其他情况下,度量方法被用于识别所需维度的数量。在社会科学和行为科学中,这点特别重要,因为研究者通常对他们观测中潜在的变异来源事先没有认识。例如,可根据一个单一的连续量,将世界上各种政治体系划分为民主系统和威权系统吗?或者说,这样做过渡简化了世界上各种社会政治结构间实际存在的一组更复杂的区别吗?

在一个度量分析中,可以通过检验一个特定的度量解的拟合优度来估计维数。实际的拟合量在各种技术之间不同。例如,对一个总加评定量表,可以用信度系数【Reliability Coefficient】估计拟合。在因子分析中,可用特征值【Eigenvalues】的数目和或共通度(communalities)作为一个总拟合(overall fit)的度量。在非标准单位的(nonmetric)多维度量中,可用应力系数(stress coefficient)估计拟合。无论具体细节如何,拟合优度通常意味着用抽象模型(由度量技术所构建的)的各组成能够预测原始数据的程度。维数通常定义为对数据仍然具有适当拟合,又有最小的维度数目的度量解。

测量

利用度量方法的第三个原因是测量。从实质的角度看,这意味着获得事物在相关变异维度上的相对位置的数值估计。根据度量模型,测量包括在由测量方法所复现的几何空间上,找到所度量的事物(仍是点或向量)的坐标。一旦实现了这一点,数值(或"量表得分")就能够作为经验变量用于后续的分析。当我们用这种方式使用度量方法时,它为本质上无法观测到的现象提供了现象的操作再现。

例如,将一组关于总统的调查问题的回答得分加总起来,可以造出一个李克特量尺,而该量尺可用于测量回答人对总统的态度。或者,一项 MDS 分析可能需要受试者评估一组职业间的相似度。在最终的 MDS 解中,点的坐标可以解释为各职业在各种认知标准(受试者用于评估职业相似度的标准)上的位置。在另一种应用中,一位研究者可能运用因子来分析州政府在一组具体公共政策中的支出。可以计算各州的因子得分,并且将其理解为各个州政府赋予各种政策领域(与不同的因子相对应)的优先性。无论情况属于哪一种,要点在于研究者假定,相较于原始变量的测量,对量表的分析能给予被考察现象更好的测量。

当度量解被用于测量的目的时,可能会面临几个问题。首先,研究者必须确认量表得分的效度【Validity】,即量表得分是他或她想要测量的属性的经验再现。这是一个棘手的问题,确切地说,因为量表经常被用来测量不可观测的现象,因此,确定两者之间的一致性可能相当困难。常用的方法是根据某些单独的、外在的标准来检验量表——这些标准或者是对同一属性的另一种测量方案,或者是其他的变量,而这些变

量同量表所测量的属性有理论上可预测的关系。如果量表得分同其他变量显示出一种经验上适当的相关模式，这就是支持研究者的解释的一个有力证据。如果预测的关系没有出现，那么在数据中，量表可能同其他来源的变化混在了一起，而这些变化源同分析者努力测量的属性是不同的。

同某些度量策略相关的另一个问题是，度量解与所考察的事物的一个数量巨大（通常是无限的），但其维度又不相同的数组（arrays）相一致。而一项因子分析或多维度量分析的最终结果即在于，在一个设定维数的空间中，确定向量或点（各自地）的位置。然而，用以界定空间的坐标轴的方向通常是任意的。如果严格地保持向量或点彼此之间的相对位置，那么坐标轴可以围绕点或向量的图形旋向其他任意位置。当然，改变一个轴的方向即改变了向量或点在那个轴上的坐标。如果向量/点的坐标具有实质意义的解释，上述旋转就是有问题的，因为随着坐标轴的旋转，量尺分的值也随之变化。就大多数度量方法而言，不存在解决这一问题的办法。因此，研究者必须根据同度量技术本身无关的因素（如简洁的解释、忠实于实质理论、与外部标准的一致），论证最终的度量解的合理性。

图形显示

实施度量分析的第四个原因需要考虑最终结果的呈现：因为度量方法试图在一个较低维度的空间内生成数据的几何模型，所以通常能够创建一个对最终的度量解的图形描述。例如，因子分析生成了一个包含向量的空间（希望用较少的维度），而空间中的向量则代表了所有被分析的变量上的共同变化。同样，多维度量可以产生一个点状图，而点之间的距离则对应于输入数据矩阵上各个事物间测量到的不相似度（dissimilarities）。此外，即使那些本质上是单维的方法（如李克特或哥特曼量表）也可以用图形显示。使用这类方法的原因之一是相比于对每个题项作图，量尺能提供所考察的属性更具体的解。此即，参照于相关的属性，一个量尺得分的直方图比任意单个题项的得分的直方图能提供多得多的关于事物分布的信息。

大多数度量分析产生的图形展示，对于梳理出分析的实质含义非常有用。解释属于研究者的一个创造性活动，它的实现有赖于研究者在由度量技术所造出的几何模型上，找出有意义的系统结构。一般来说，通过对统计结果的视觉检视达到这个目的很方便，因为相比较于数字化的信息，模式化的规律通常更容易在图形中被识别出来。

结论

总之，度量方法论提供了可用以模型数据矩阵中存在的系统性结构的相应策略。如果这些结构被认为可归之于数据中存在的某些潜在的，或不能直接观测到的变化源，这类方法就尤其适用。后一项特征恰恰是度量在社会科学和行为科学中如此重要的原因：实质理论中的核心概念本质上是不能直接观测到的；而度量方法为概念的操作化提供了工具。然而，量尺仅能生成模型，它们本身并无所谓“对”或“错”。它们仅仅代表了分析者关于某些过程的性质的理论，而正是那些过程产生出经验观测。度量分析可能产生相关的理论洞见和有用的经验变量。但是如同其他所有理论，它们还只是对实在的性质的试验性表述，因此，如果针对同一数据，一个不同的度量模型能生成一个具有同等程度

的间省、解释力,以及分析上有效的几何表示,原模型的有效性即受到质疑。

——William G. Jacoby

(马妍译　赵锋校)

参考文献

Coombs, C. H. (1964). *A theory of data*. New York: Wiley.

Jacoby, W. G. (1991). *Data theory and dimensional analysis*. Newbury Park, CA: Sage.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.

Young, F. W., & Hamer, R. M. (1987). *Multidimensional scaling: History, theory, and applications*. Hillsdale, NJ: Lawrence Erlbaum.

散点图(Scatterplot)

散点图是展示两个定量变量【Quantitative Variable】之间关系【Relationship】的二维坐标图,即将一个数据集中的每个观测值都表示为图上的一个点。尽管在正式的出版物中,它们的使用频率比其他统计图低,但可以有理由认为散点图是数据分析中最重要的图。散点图不仅可以直接画出变量的变化,还能用于检验导出的量,如一个回归模型【Regression Model】中的残差【Residual】。

散点图的先驱者至少可以追溯到 17 世纪,诸如费马和笛卡尔对坐标平面的发展,以及伽利略对运动的研究。散点图的第一次统计学运用可能来自 19 世纪的高顿爵士,同时,散点图与他对相关【Correlation】、回归和双变量正态分布【Normal Distribution】的前沿工作紧密联系在一起。

图 1 中给出了一个示例性散点图的若干样式,展示了 50 个发展中国家中生育率同避孕品使用间的关系。总和生育率是一位存活过其开始生育年之后的妇女在当前年龄段的生育水平上期望的活产婴儿数。根据惯例,解释变量(避孕品使用)出现在横轴,而响应变量(生育率)出现在纵轴。

①图 1(a)显示了传统的散点图。

②在图 1(b)中,国家的名字显示在数据点旁边。为了最小化重叠现象,名字以随机角度出现,但是许多名字仍然难以识别。如果放大图像,但不按照比例同时放大图上的文字,这个问题能得到一定的缓解。

③图 1(c)展示了更现代式的散点图:取消了网格线;刻度线不再放入图中;数据被绘制为空心圆而不是实心圆,以便更容易辨别部分重合的点;坐标轴画的只是围住数据。图中的实曲线代表一个稳健的【Robust】非参数回归【Nonparametric Regression】平滑;两根虚线是根据非参数回归的正残差和负残差作出的平滑,也是在每个方向上残差的标准差【Standard Deviation】的稳健估计。生育率和避孕品使用之间的关系几乎是线性的【Linear】;残差具有接近常数的散布,并且围绕回归线呈对称分布。关于平滑散点图的更多信息可以在福克斯(Fox, 2000)的文章中找到。

④图 1(d)展示了如何将一个第三个类别变量的值编码在散点图上。图中用不同的字母对地区作了编码。相似地,还可以使用不同的颜色或图示符号进行编码。

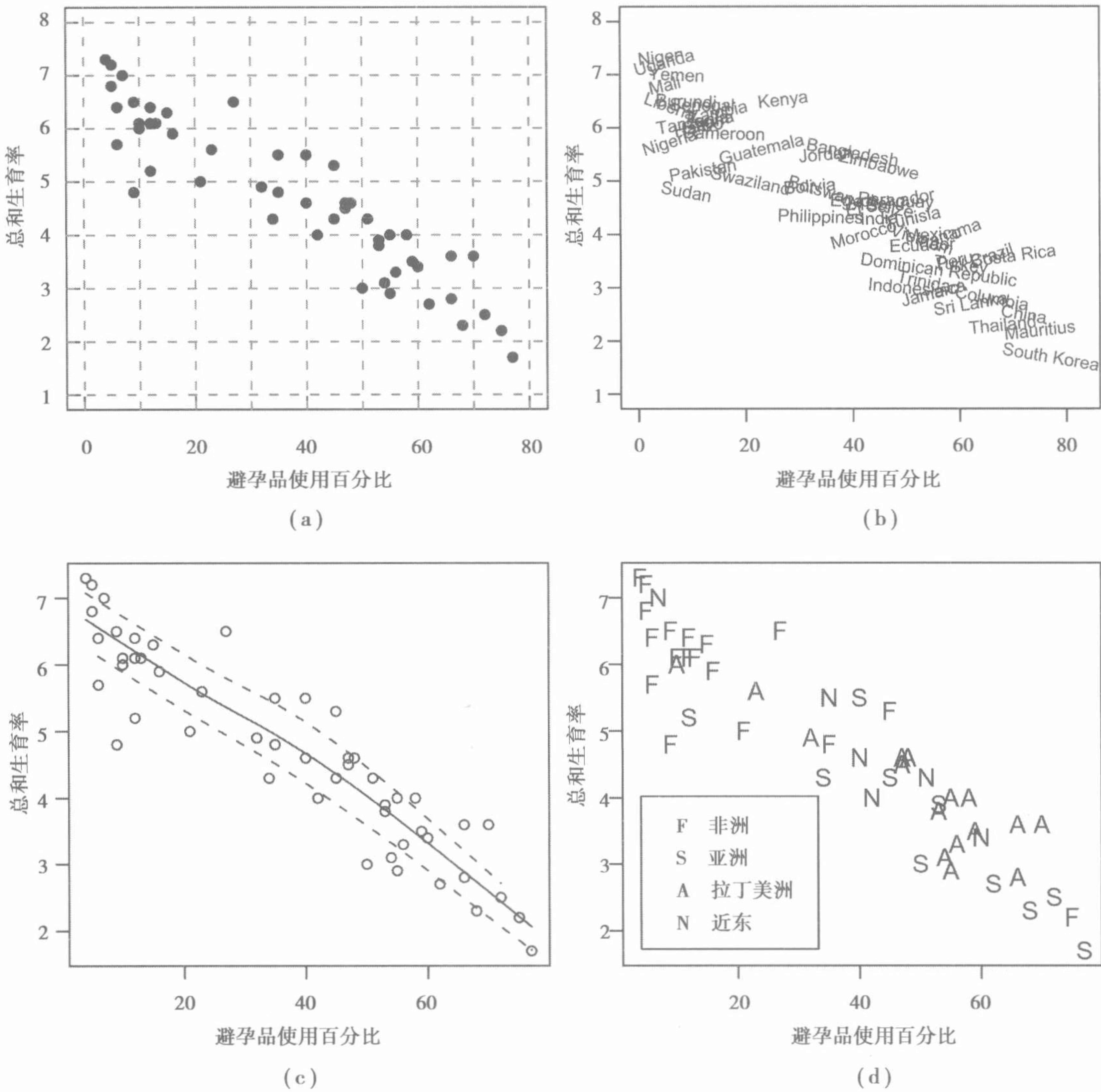


图 1 4 种样式的总和生育率(每个妇女的期望生育数)对育龄期已婚妇女避孕品使用百分比的散点图,50 个发展中国家

除了图 1 中的例子,简单的二变量散点图还有许多重要的变化样式和扩展样式:三维动态散点图可以在计算机屏幕上旋转,以检验 3 个变量之间的关系;散点图矩阵可以显示一组变量中所有成对的边缘关系;而交互式统计图可以用来识别散点图中的点,并可以将散点图同另一个其他的散点图,以及其他的图形展示联系在一起。这些内容以及其他的方法在克利夫兰(Cleveland,1993)和雅各比(Jacoby,1997)的书中有讨论。

——John Fox

(马妍译　赵锋校)

参考文献

Cleveland, W. S. (1993). *Visualizing data*. Summit, NJ: Hobart.

Fox, J. (2000). *Nonparametric simple regression: Smoothing scatterplots*. Thousand Oaks, CA: Sage.

Jacoby, W. G. (1997). *Statistical graphics for univariate and bivariate data*. Thousand Oaks, CA: Sage.

Robey, B., Rutstein, S. O., & Morris, L. (1992). The reproductive revolution: New survey findings (Technical Report M-11). *Population Reports*. Baltimore, MD: Johns Hopkins University Center for Communication Programs.

Scheffé 检验(Scheffé's Test)

亨利·谢菲(Scheffé,1953)发展了该检验,其目的在于确定在一个方差分析【Analysis of Variance】中(其中各个组的得分并不相关),哪些均值或均值的组合显著地不同于另一些均值或均值的组合,而其检验条件在于,已知在此方差分析中,对这些均值或均值组合的总的 F 比率【F Ratio】检验是显著的,但是在分析前,对于那些均值或均值组合间具有显著差异并无预测。这类检验通常被称为事后(post hoc)或者后验(posteriori test)检验,或者是多重比较检验(multiple comparison test)。它考虑到一个事实,即做得比较越多,一个显著的差异就越可能出现。它比较适宜对规模不等的组进行比较。与其他事后检验相比,由于更不可能基于机遇获得一个显著的差异,它是一种更加保守的检验(Toothaker, 1991)。

在方差分析中,若两个以上组别的某个因子或两个及更多因子间的交互效应具有显著的 F 比值,这只能说明组间的方差显著地不同于基于机会而预期的方差,但不能说明哪些均值或均值组合之间有显著的不同。如果存在着预测特定均值或均值组合间差值的强有力的理由,那么可以用一个先验检验来确定这些差值的统计显著性,如用一个单尾的 t 检验【t-Test】。如果不存在预测特定差值的强有力的理由,通常可能作出的比较的数目就会更多。例如,可用下面的公式计算,均值两两相比的总数:$[J\times(J-1)]/2$,其中 J 是均值的数量。因此,如果有 4 个均值,可能的配对比较总数是 6,即$[4\times(4-1)]/2=6$ 次。如果发现一个显著差异的概率被设定为通常的 0.05 的水平,那么发现一对或更多的比较对在这一水平上显著的概率,可由公式 $1-(1-0.05)^J$ 算得。这个概率水平被称为总推断错误率(familywise error rate),如果进行 4 次比较,这个值大约是 0.17,即 $[1-(1-0.05)^4]=0.17$。然而,除了均值的配对比较,还存在着其他的比较,而这些比较必然增加潜在比较的总数目。例如,可以将第一和第二组均值合并在一起,同第三组或第四组的均值,或同第三和第四组均值的组合相比较。

Scheffé 检验通过用一个因子或交互效应的 F 比值的临界值乘以减 1 的被比较群组数目,来控制包括配对比较以外的其他比较的总推断错误率。假如有 4 个组,每组有 5 个个案,在 0.05 的显著水平下 F 值的临界值是 3.24,其分子的自由度为 3,分母的自由度为 16。因此,对这 4 组均值的任意比较而言,Scheffé 检验的 F 比值的临界值为9.72[$(4-1)\times3.24=9.72$]。如果方差分析的 F 比值仍然显著,那么,在所有均值或均值组合的比较中至少有一组必然是统计显著的。

——Duncan Cramer

(马妍译 赵锋校)

参考文献

Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika*, *40*, 87-104.

Toothaker, L. E. (1991). *Multiple comparisons for researchers*. Newbury Park, CA: Sage.

碎石图(Scree Plot)

雷蒙德·卡特利(Cattell,1996)开发出了碎石检验,来代替如 Kaiser-Guttman 一类的检验,用以确定一个因子分析【Factor Analysis】中的因子数目,即那些能够解释在多个变量间共享的或共有的某个具有一定量级的方差的因子数目。这一检验方法的首要观点在于,一个因子解释掉的方差量似乎与后续的一系列因子解释掉的方差量非常相似,然而,这些量一般而言会逐渐变小,由此表明这些后续因子是可以忽略的。碎石是一个地质术语,用来描述出现在斜坡地基中的瓦砾和石块,这些碎石掩盖了斜坡真正的地基。在这个意义上,这一检验常用于确定哪些因子可以被视为碎石或碎片,亦即哪些因子对变量的任何潜在结构没有实质性的贡献。

在碎石检验中,由因子解释掉的方差的量被做成如图 1 一样的图。图的纵轴表示一个因子解释掉的方差量。横轴按照因子解释掉的方差量不断减少的方式,给出了该因子的序号也就是该因子出现的顺序。一个因子所解释掉的方差量常被称作特征值【Eigenvalues】或特征根(latent root)。特征值是每个变量与那个因子间的因子负荷的平方和或每个变量与那个因子间的相关系数的平方和。图 1 显示了对 13 个变量的一个主成分分析的特征值,该分析得到了 13 个因子或成分。碎石看上去出现在成分 3。换句话说,检验建议 13 个因子中有 2 个能被保留。碎石开始的点可以由一条直线确定,该线穿过或非常接近碎石所代表的各个点,如图 1 所示。位于该线之上的因子代表保留下的因子数目。

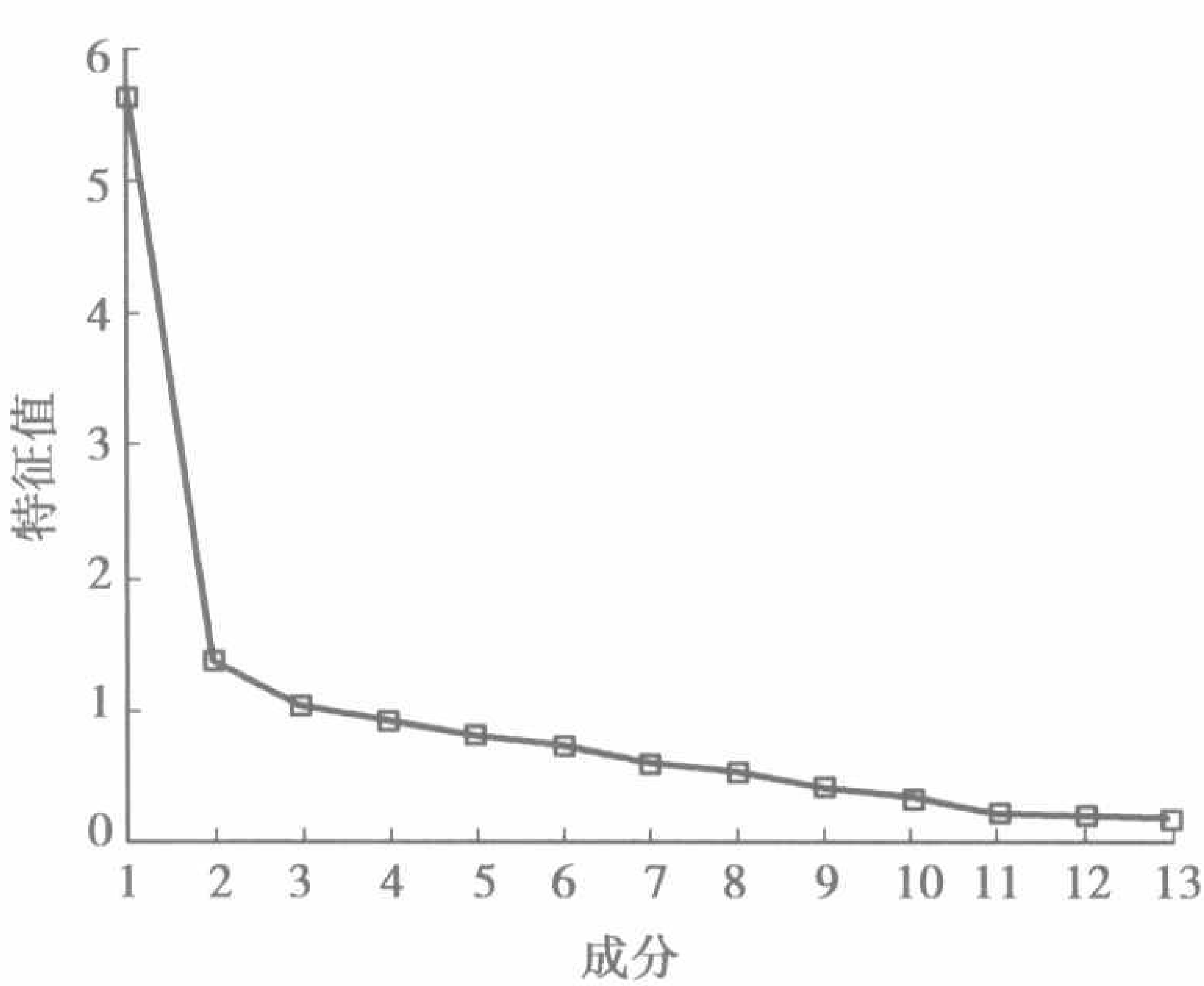


图 1 碎石图

一个因子分析或主成分分析总会产生与变量数同样多的因子数或成分数。然而,因子分析的目标在于确定诸变量是否能够组合在一起,从而构成一组数目较少的可称为因子(factors)的聚合变量或超变量。在诸因子中,第一个因子总会解释掉最大数量的方差。后续因子所解释掉的方差量会逐渐变小。这就需要建立一条标准用以确定多少后续的因子是可以忽略的,因为它们解释掉的方差太小。Kaiser-Guttman 检验判定如果因子的特征值为 1 或小于 1,该因子当被忽略,因为该因子至多解释了一个单一变量所能解释掉的方差量。以卡特利的经验而言,这项标准会导致变量少则因子也少,而变量多因子也多,虽然他没有具体说明这些形容词指代的数字是多少。兹维克和韦利瑟(Zwick & Velicer,1986)提出当变量数目为 36~72 时,该标准会高估因子数目,其估出的因子数目会远多于碎石检验给出的因子数目。

——Duncan Cramer

(马妍译 赵锋校)

参考文献

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245-276.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*, 432-442.

定性资料的二次分析(Secondary Analysis of Qualitative Data)

定性资料的二次分析,也称作定性二次分析(qualitative secondary analysis),是为了探究与原来的质询不同的研究问题,再次考查一个或多个即存的由定性方式取得的资料集。定性质询常涉及使用诸如半结构化访谈【Semistructured Interview】、参与观察【Participant Observation】及田野工作【Field】方式收集到的大量资料集,而以这些方法产生的资料集通常包含有相当丰富的信息,以致很难将它们都包含在初次的报告中。而且,随着时机的成熟,新出现的问题常会将即存的资料集变成产生切合实际的知识的一个有效且适当的来源。

二次分析也提供了一种机制,透过这种机制,在相似的田野条件下工作的研究者就可以将不同来源的资料结合成更大的资料集,并以此去回答某个比较性质的问题或考察那些在小样本背景下意义模糊的主题。当得自不同人群、地理区域、时期或情境条件的资料集适于二次质询时,二次质询就能以说明原质询方法的隐含意义的方式,更准确地呈现出感兴趣的现象。例如,安戈斯特和迪特里克(Angst & Deatrick, 1996)对比了各自先前对患有囊性纤维化的儿童生活的研究和对那些将接受脊柱侧弯手术的儿童生活的研究,并通过这种比较的二次分析,引发我们注意到患慢性病的儿童的日常生活与重大的治疗决策之间的矛盾。

虽然定性研究【Qualitative Research】者过去常在一个资料集的基础上写就多个报告,但是明确地将二次分析作为一个独特的定性研究方法则是相对近期的现象(Heaton, 1998)。由于二次分析的目的在于澄清理论的、方法的参照框架——定性研究者以此进行其初始研究,因此方法论的精确和透明对一个二次分析在认识论上的一致性和可靠性就至关重要(Thorne, 1994)。无论资料集的数量有多少,也无论二次研究者可能在构造资料集时的涉入水平有多深,一个二次分析本身包含了用以理解抽样、资料汇集程序及资料汇集与分析间交互的特有方法(Hinds, Vogel, & Clarke-Steffen, 1997)。

定性二次分析可能以一种富有效率和效果的方式提出那些超出任何单个研究团体所能解决的问题,考察那些在初次质询时没有考虑的现象,或者探索那些仅当处于同一研究领域的研究者互相考查其研究发现时才会显出理论潜力的研究主题。

——Sally E. Thorne

(王玥译 赵锋校)

参考文献

Angst, D. B., & Deatrick, J. A. (1996). Involvement in health care decisions: Parents and children with chronic illness. *Journal of Family Nursing*, *2*(2), 174-194.

Heaton, J. (1998). Secondary analysis of qualitative data [Online]. *Social Research Update*, *22*.

Retrieved April 28, 2002, from http://www.soc.surrey.ac.uk/sru/SRU22.html.
Hinds, P. S., Vogel, R. J., & Clarke-Steffen, L. (1997). The pos-sibilities and pitfalls of doing a secondary analysis of a qualitative data set. *Qualitative Health Research*, 7, 408-424.
Thorne, S. (1994). Secondary analysis in qualitative research: Issues and implications. In J. Morse (Ed.), *Critical issues in qualitative research methods* (pp. 263-279). Thousand Oaks, CA: Sage.

定量数据的二次分析(Secondary Analysis of Quantitative Data)

二次分析指为了不同的目的,分析由其他研究者收集的原始数据。本词条集中讨论对来自抽样调查【Survey】和人口普查【Census】的定量数据的二次分析。有关二次分析的各种界定及其广泛的数据源,可参见海曼(Hyman,1972)、哈基姆(Hakim,1982),以及戴尔等人(Dale, Arber & Procter,1988)的讨论。

在经常受有限资金限制的研究者那里,高质量数据的二次分析可以成为一种极具经济效率的工具,即可以用很少的额外花费获得很大的收益。完全避免了去计划和实施一次抽样调查,以及清理和存档数据这类既耗时又费钱的过程。然而,由于数据是由其他研究者收集的,二次分析者就没有机会去影响调查中所提的问题或所用的编码框【Coding Frames】,因此,分析者也需要在分析的全过程将这一重要的因素铭记在心。

历史发展

随着计算机在1950年代和1960年代的发展,定量数据的二次分析成为可能。计算机化意味着数据可以以压缩的方式储存,且方便移动,而最重要的是,可以对数据作匿名处理,从而在不泄露数据源身份的前提下由他人使用。用于统计分析的计算机软件包,如SPSS,使得更多的人能以更方便且更复杂的方式分析调查数据,而在此之前这几乎是不可能的。

作为对这些发展的回应,以机器可读的数据的积累、归类、储存和散布为目的的数据库【Data Archives】得以建立。1960年代,美国建起了一些数据库如罗珀舆情研究中心(Roper Public Opinion Research Center)、位于密歇根大学的政治和社会研究校际协会(the Inter-university Consortium for Political and Social Research, University of Michigan)。欧洲和英国也建起了各自的数据库,如位于艾塞克斯大学的英联邦数据库(UK Data Archives, University of Essex)。从这开始,使用者的数量和累积的研究数量都在稳步增长。此外,数据类型也在扩展,不仅涵盖了行政记录,还汇集了诸如来自地区水平的普查统计数据。

应用

在英国和美国,大规模政府调查数据的可得性是另一个对二次分析的发展有相当重要性的因素。政府调查可以提供带有自身特性,而难以由其他途径获取的数据。这些调查通常具有全国的代表性;有很大的样本量;通过人际访谈来获得资料;同时,最重要的是,由那些在抽样【Sampling】、问卷设计、田野工作【Field】,以及抽样调查工作的所有其他方面都相当专业的固定职员来实现。最终结果是,这些调查可以提供涵盖更广、质量更高的数据,要超过那些大多数学者都可以申请到的基金所能提供的。在下述领域,二次分析有其特殊价值。

针对广泛主题的全国层面的概推

二次分析能够提供一个基础，在此基础上，研究者可以将全体居民【Population】作为整体作出概推。这时，大规模的政府调查就特别重要，因为它们通常以相当的深度覆盖范围广泛的主题，而且拥有大的足以给出很高精度估计的样本。

历史比较

二次分析可能是可以作历史比较的唯一手段，特别是当信息不能通过回想的方式来收集时，例如，收入和花费，或态度。数据库容许研究者回到过去，就人们想了些什么、如何投票、挣了多少钱等问题，也许是20年前的情况，找到相应的信息资源。许多政府调查始于1970年代或更早，从而为历史的变迁分析提供了条件，例如，过去20年间女性在家庭中的经济角色。

纵贯分析

人们逐渐认识到纵贯数据对检验假设【Hypotheses】和确立因果【Causal】过程的价值。然而，对大多数研究者而言，由于建立和维护一个长期的纵贯研究困难，二次分析就是他们的唯一选择。在英国，许多纵贯研究［如全英家庭追踪调查（the British Household Panel Survey）］就是作为国家资源而专门设计的。在美国，如收入变迁的追踪研究（the Panel Study of Income Dynamics）和全国青年的纵贯研究（the National Longitudinal Study of Youth）一类的纵贯研究都为学术研究提供了类似的丰富资源。

国际比较

国际比较虽然极具研究的重要性，但是带来了大量费用和组织方面的问题，这就给二次分析提供了另一个可以发挥关键作用的领地。有时，研究者可以就某些主题和问题，找到具有充分相似性的源自不同国家的数据，以便进行比较研究，例如，英国、法国和瑞典间的社会流动研究（Erikson, Goldthorpe, & Portocarero, 1979）。还有些时候，抽样调查的设计就是为了国际比较，例如，欧洲社会调查（European Social Survey）的设计就是为了量化地描绘和解释正在变化的欧洲各成员单位间的互动，以及各成员单位的居民态度、信仰和行为模式。

小数群体的代表性数据

二次分析同样为那些缺乏明确的抽样框【Sampling Frame】的少数群体提供一种获取数据的方式。许多人口调查都拥有一个相当大的样本，从而为相对小的子群体的分析提供了条件。例如，一项对老年群体的研究，为了确定一个合适的样本，可能需要进行相当严格的筛选，而这可以通过二次分析轻松地实现。来自普查的微数据（microdata）样本通常也大得足以支持对少数族裔群体的分析（Dale, Fieldhouse, & Holdsworth, 2000）。

作为定性研究的背景数据

把一个有全国代表性的维度纳入一个定性研究，以便确立研究现象的规模和分布，这样做通常很有价值。例如，一项关于单亲父母的开支决策的定性研究，通过将他们的背景信息，如单亲父母的居住地，以及他们在整个居民中所处的收入和开支水平包括到研究中，可能得到相当程度的品质提升。通常，经由二次分析获取这种背景信息既廉价又相当方便，而在同等条件下，一项具有全国代表性的抽样调查的成本就大得不用考虑了。应用二次分析来辨识兴趣群

体的社会和人口统计特征也会有助于定性研究者选择所要调查的样本。

反过来说,二次分析可以突显那些需要通过定性方法去求解的令人感兴趣的关系。

数据的保密

尽管数据的保密【Confidentiality】是所有社会研究的一项基本考量,但是,当为某一特定目的收集的数据可以为其他人的分析所用时,这点就尤为重要。因此,必须采取一些标准的预防措施来确保个人身份不被识别,其中包括对数据作匿名处理,即勿使数据带有调查对象的姓名和地址,将数据中的地理区位信息限制在不能辨识个人身份的水平上,以及消除数据中使得某个回答人成为唯一可辨识的任何特别或不一般的特征。

——Angela Dale

(王玥译　赵锋校)

参考文献

Dale, A., Arber, S., & Procter, P. (1988). *Doing secondary analysis*. London: Unwin Hyman.

Dale, A., Fieldhouse, E., & Holdsworth, C. (2000). *Analyzing census microdata*. London: Arnold.

Erikson, R., Goldthorpe, J. H., & Portocarero, L. (1979). Inter-generational mobility in three Western European societies. *British Journal of Sociology*, *30*, 415-441.

Hakim, C. (1982). *Secondary analysis of social research*. London: George Allen and Unwin.

Hyman, H. H. (1972). *Secondary analysis of sample surveys*. Glencoe, IL: Free Press.

调查数据的二次分析(Secondary Analysis of Survey Data)

调查数据的二次分析是以不同于原始研究问题的新研究问题为目的,对现存的抽样调查回答的再分析。随着可为研究者所用的数据集的不断增加,在社会研究中,抽样调查的作用已然扩大,同时,随着数据质量的提高和获取机会的增多,数据的二次分析也被赋予日益中心的地位。二手调查数据可用于诸多主题,包括健康、态度、人口特征、犯罪行为、老龄化等,而且许多现成的数据集既可用于截面分析,也可用于时序分析。二次调查分析的一个主要优势是节省资源。既节约了时间和金钱,也不需要大量人手。因为数据已经存在,所以分析通常不会对调查对象有所伤害,尽管相关讨论,即二次分析对人类主体的牵连,还在持续进行。因为有多组数据集可用,这就简化了比较分析——通常有国家间的比较或历时的比较。

这个方法的局限通常在于效度【Validity】的问题,即现成数据用于特定研究问题的适当性。因为二手研究者并不设计原始调查,所以数据集中未必包含所有感兴趣的变量。即便有所有的变量,它们也可能是以不那么有效的方式测量的。使用现成的数据,研究者必须确保样本代表了目标总体,需要研究的子样本包含充足的个案,且容易提取。分析单位【Unit of Analysis】也必须与研究相一致。尽管许多现存的数据有特定的关注,人们已付出了许多努力,以促进综合调查,即涵盖有广泛主题的调查。最广为人知的是由芝加哥大学的国家舆情研究中心(the National Opinion Research Center,NORC)开展的综合社会调查,含有一系列的社会指标数据。

二次调查数据现在多数存储于硬盘、光碟和网络上，便于取用。人们还设立了许多数据库，其中一些数据库收集了大量数据，涵盖众多的主题，另一些则小而专门。不仅学术资料库、政府机构和民意研究中心（例如罗珀和盖洛普）是二次调查数据的便利来源，任何存有此类数据的组织（如私有研究公司、基金会、慈善组织）都可视为二次分析者的数据来源。密歇根大学的政治与社会研究的大学间协会（the Inter-university Consortium for Political and Social Research，ICPSR）是学术性的社会科学研究者的一个主要数据来源，可以从相关网址进入。ICPSR 是寻找所需数据好的起点。除了它们自有的大量数据，该网站给出了其他数据库的链接。

为了最大限度地使用现成的调查数据，研究者应该把一般的实质研究兴趣与对现成数据的熟悉融合起来，能够根据数据的调查设计和实施、抽样的问题，以及数据库的局限来评估主要的数据来源。研究者还必须以他们富于创造性的方法去合并一个调查或多个调查中的数据，以及通过整合来自其他数据集的数据来合并数据。尽管使用现成的调查数据总会遇到内在的挑战，但是对社会研究而言，机遇同样巨大。

——Laura Nathan

（王玥译　赵锋校）

二手数据（Secondary Data）

分析者可通过实施抽样调查【Survey】、做实验【Experiment】、对新闻材料进行内容分析【Content Analysis】，或者对观察到的行为进行编码【Coding】等方式来收集数据，并且所有这些方法产生的都是原始数据。相反，二手数据由他人收集，而非分析者自己。社会科学工作的相当部分是处理二手数据。例如，一个社会研究者可能去图书馆收集由美国人口普查局最初收集整理的数据。或者，一个心理学家为了开展某一研究的元分析【Meta-Analysis】，可能收集大量已发表的检验某一特定假设的实验数据。另一个例子是使用存于数据库中的数据，例如政治和社会研究大学间联盟这一数据库。越来越多的二手数据可以通过网络获取。二手数据的优势在于它们的成本比原始数据小很多，因为你不需要自己去收集。它们的劣势在于二手数据可能不包括特定的观测【Observation】或变量【Variable】，而你会将这些观测或变量纳入你实施的研究中。越来越多的情况是，综合社会抽样调查（例如美国综合社会抽样调查）在每一次的调查中并没有一个特定的主研究目标，其目的则在于便于研究者将其作为“二手”的原始数据予以应用。

——Michael S. Lewis-Beck

（王玥译　赵锋校）

二阶（Second-Order）

参见阶【Order】。

相依回归(Seemingly Unrelated Regression)

当两个或多个拥有不同自变量【Independent Variable】集的回归【Regression】方程,通过相关的扰动项【Disturbance Term】而彼此发生关联时,可以使用相依回归或相依回归方程(seemingly unrelated regression equations,SURE)。特别是当各回归方程的扰动项高度相关,而各自变量集不存在高度相关时,用 SURE 进行估计会比用单独的回归估计得出更有效的系数。SURE 估计方法可用于集合的(pooled)时间序列截面【Time-Series Cross-Section】数据,而此类数据或者受时间序列的支配或者受截面的支配。它也可单独用于横截面数据【Cross-Sectional Data】和时间序列数据【Time-Series Data】。如果方程的误差项不相关,那么可用普通最小二乘【Ordinary Least Squares,OLS】回归处理各方程。

一个关于联邦机构预算的模型可以说明 SURE 方法的应用。如果因变量【Dependent Variable】是各联邦机构的支出水平,同时,若根据不同的机构类型设立各个模型,那么各个模型的扰动项很可能相关。此即,不同类型机构的支出水平很可能都受到共同的外生约束的影响。例如,美国对恐怖主义的战争会产生对预算的大量需求,从而影响了各非防御机构的支出水平。

当方程组中的每个方程有不同的自变量集时,SURE 也是最佳选择。在预算的案例中,这能让研究者认识到并非所有的机构都是类似的,从而根据类型将各机构归组到不同的模型中,并且按类型将机构分组为不同的模型。按类型将各机构归组,并将之置入各个方程中,这样的做法容许各机构既受一组相同自变量的不同影响,也受不同自变量的影响。

例如,失业对诸如核监管机构这类监管机构预算的影响,不同于对劳动部下属的就业和培训司预算的影响。而失业对于联邦通信委员会这样的监管机构的预算可能完全没有影响,这就意味着不要将这类机构放在模型中。与之相反,对集合机构数据的 OLS 或广义最小二乘【Generalized Least Squares,GLS】估计假定所有的机构都以同样的方式受到同一组自变量的影响。这样做就不仅将数据集合在一起,还将估计值都集合在一起了。

OLS 或者 GLS 对混合机构数据的估计假定所有的机构都以同样的方式受到同样的自变量影响,也就是说,不仅数据是混合的,估计的结果也混为一谈。

——Charles Tien

(马妍译　赵锋校)

参考文献

Kmenta, J. (1986). *Elements of econometrics* (2nd ed.). New York: Macmillan.

Pindyck, R., & Rubinfeld, D. (1991). *Econometric models and economic forecasts* (3rd ed.). New York: McGraw-Hill.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *American Statistical Association Journal*, *57*, 348-368.

选择性偏倚(Selection Bias)

选择性偏倚是任何社会科学研究设计的一个重要的影响因素,因为它的出现一般会导致不准确的估计。当样本的观测值依赖于目标变量的取值时,就会出现选择性偏倚。这时,样本就不再是从研究总体中随机抽取的,因此,根据这一样本所做的总体推论就是有偏倚的。尽管研究者应该通过细致的研究设计来减少非随机的选样,但是许多情况下,这一问题难以避免。当数据的生成过程不由研究者本人控制时,问题尤其严重。由此,研究者发展出一些方法试图修正与选择性偏倚相关的问题。总体来说,这些方法包括将选择过程模型化,从而在评估结果变量时控制这个过程。

举例说明选择对推论【Inference】的影响。请考察 SAT 分数对大学成绩的影响。如果假设较高的 SAT 分数相应于大学的成绩,人们就期望能够发现,在被最好的大学录取的学生中,两者间(SAT 分数和大学成绩)存在着强的正向关系。然而,以低的 SAT 分数而被录取的学生必然有其他一些优势才能被录取。因此,与那些 SAT 分数相同,但没被录取的学生相比,这部分学生会取得非常高的成绩。这就使得 SAT 分数对大学成绩的估计发生偏倚,即认为影响趋于零。要注意,如果样本中的大部分学生都有高的 SAT 成绩,就不会发生选择性偏倚;但是,因为样本中有低 SAT 分数的学生不能代表所有低 SAT 分数的学生,且在所有低 SAT 分数的学生中,他们在一些与大学成绩相关但未被测量的因素中肯定有较高的得分,所以就会发生选择性偏倚。

关于选择性偏倚的一个常见的误解是,认为如果只试图得出所分析的样本的推论,而不试图推论目标总体,就可以不考虑选择性偏倚,类似地说,正确地估计所有被录取者的 SAT 分数和大学成绩之间的关系。这个想法是错误的:因为在我们的例子中,样本学生的大学成绩既不代表所有的大学申请者,也不代表所有有相同 SAT 分数的学生。换言之,我们的样本不是一个以所有学生的 SAT 分数为条件的随机样本。因此,估出的关系是不准确的。这个例子中,由于选择过程,回归系数会被偏估而趋于零。然而,一般而言,偏倚可能趋向任意方向,此外,在更加复杂的(多变量)回归中,将无法获知偏倚的方向。

尽管选择性偏倚可以有不同的形式,但是它一般可以分为两种不同的类型:截删和删节【Censoring and Truncation】数据。截删之所以发生,是当每一个被观测对象的所有特征均被观测了,但是目标因变量有缺失。删节之所以发生,是因为当某些个体由于被屏蔽于样本之外时,他们的特征没有被观测到,于是所有的观测值不是完全存在就是完全缺失。一般而言,截删更容易处理,因为有数据可用于直接估计选择过程。处理删节的统计模型则更多依赖于相应的假设,即选择过程同影响结果变量的过程是如何发生联系的。

最著名的用于自选择过程(self-selectivity)的模型是詹姆斯·赫克曼的估计量【Estimator】,该估计量用于一个连续的、存在截删问题的目标因变量(Heckman,1979)。该估计量可以修正线性回归背景下的选择性偏倚,已广泛地用于社会科学之中。赫克曼的修正尝试通过使用一个 Probit 方程来模型化选择过程,即将自变量的观测值是否被截删当作指标,进而估计选择偏倚。然后,估计目标方程的误差项的条件期望值,因为,如果目标方程的误差项同选择方程的误差项相关,条件期望值就不为零。

一个被称为迷 Mills 比率(inverse Mills' ratio)的变量会被作为额外的回归项带入目标方程,并用 OLS 法做估计。如果以此方式估计模型,就需要调整标准误,因为模型的标准误低估了模型中的不确定性。另一种方式是,对于两个方程同时用最大似然技术予以估计,并无须调整标准误。在满足假设的情况下,赫克曼估计量给出了方程的无偏且一致的估计值,同时,也给出了两个方程误差项间的相关估计值。

在赫克曼模型中,通过对选择项系数的显著性检验,或者,更一般地,通过似然比检验【Likelihood Ratio Test】,即考察选择模型相对于无关联模型的拟合度,可以检验是否存在选择性偏倚。在解释自变量对目标因变量的效应时必须谨慎,因为某些因素既可通过选择方程间接地发生影响,也可能直接地影响目标方程。另一个重要的考量是,是否存在相应的因素,它们影响选择方程,但不直接地影响目标方程。尽管在没有这些因素的情况下,赫克曼模型是可判定的(仅就函数形式而言),但是,将这些仅属于选择方程而同目标方程无关的因素加入模型,通常会增进模型的信度,因此,它们能够增加迷 Mills 比率的变异度。

针对更复杂的情况,发展出了许多高级的统计技术,其中包括不可能直接估计选择方程的被删节数据,以及目标因变量是两分的,而不是持续的数据(Dubin & Rivers, 1989)。对这些方法的概要介绍,可参见马达拉的著作(Maddala, 1983);更高级的方法包括放松分布假设的半参数方法,可参见温希普和梅尔的论文(Winship & Mare, 1992)。总之,由于发现标准的选择性偏倚模型常对修正的假设(assumptions of the correction)很敏感,特别是关于两个误差项呈双变量正态分布的假设很敏感,统计专家已经发展出了许多高级的方法。

——Frederick J. Boehmke

(王玥译　赵锋校)

参考文献

Dubin, J. A., & Rivers, D. (1989). Selection bias in linear regression, logit and probit models. *Sociological Methods & Research*, *18*, 360-390.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*, 153-161.

Maddala, G. G. (1983). *Limited dependent and qualitative variables in econometrics*. Cambridge, UK: Cambridge University Press.

Winship, C., & Mare, R. D. (1992). Models for sample selection bias. *Annual Review of Sociology*, *18*, 327-350.

自填式问卷(Self-Administered Questionnaire)

自填式问卷是研究中最常用的收集数据【Data】的方法之一。此外,自填式问卷还出现在我们许多的生活场景中。比如,从幼儿园到研究院一直都在用的大多数课堂测验。课堂测验是一种自填式问卷。相似地,为了获取驾照及死亡证明这类东西,我们也需要填写各种表格或问卷。

自填式问卷之所以在日常生活的各个场景中大量出现,并为人所熟悉,其常见的原因在于,人们自以为可以一夜之间就设计出一份真正的问卷,可以立即用其收集数据。但是,如同任何一项研究努力,如果研究者想要最大限度地收集到完整、可信和有效的数据,就需要在自填式问卷的设计上花费时间和投入思考。

自填式问卷的类型

问卷既可以完成于有督导的环境，也可完成于无督导的环境。在有督导环境下，有人负责回答填答人可能会有的问题，监督填答人在填写问卷时是否相互交谈，以及督促填答人完成问卷，并将问卷交回研究者或其他人。在课堂上分发问卷是有督导的群体填写的一个例子。问卷也可以在结构松散的环境中发给个人，如注册登记的等候队列，这时，虽然有数据收集者，但并不明显。

邮寄给回答人的问卷是最常见的无督导填写的例子，不过，越来越多的问卷通过电子邮件、互联网和其他在线方法【Online Research Methods】进行填写。无论处于何种环境，自填式问卷都必须独立使用，或完全是自说明的，当使用无督导的填写程序时，这点尤为重要。

优点和弱点

自填式问卷的一个最大优点是，相对于其他数据收集方法，它的成本较低。邮寄问卷有三个与样本相关的优点，即更广的地理覆盖、更大的样本，以及对某个样本总体的更大的覆盖率。同时，所有自填式问卷都更容易实施，因为数据收集、操作和分析需要的人力更少。不像大多数其他数据收集方法，当问卷通过邮寄或网络分发时，可以假定调查样本中的所有成员大致能同时收到问卷。许多研究者认为，就敏感话题【Sensitive Topics】而言，通过自填式问卷可以获得更完整且更可信的信息。

邮寄和自填式问卷也有很多的弱点，而这些弱点会限制或阻碍它在许多研究环境中的使用。如果使用邮寄或网络分发问卷，就必须有可用的地址清单。邮寄问卷的一个最大且被研究最多的弱点是低回答率。如果没有努力通过额外的邮件或电话联系来追踪原初的邮件，那么一个普通的社区样本的预期回答率在20%左右。目前，网络调查回答率似乎更低。如果总体的20%据推测是未受教育的文盲，或者总体中有相当比例的回答人不能阅读研究使用的语言，就不能使用自填式问卷。老年人或其他有视力问题的人群也可能有填答问卷的困难，或根本无法填写。

只有当研究目标清晰且不复杂时，才可使用自填式问卷。一个清晰且不复杂的数据收集目标需要就如何制作问卷，以及防止使用那些常用于访谈中的许多典型的问卷设计策略，注意以下诸点。第一，问卷必须简短，因此必须减少所提问题的数目和所收集数据的总量。一般认为一份自填式问卷的上限不超过12页，但是长度需视回答人的动机而变。如果话题对回答人而言特别重要，他们可以容忍较长的问卷。第二，问题必须是封闭式【Closed Ended】的。尽管有些动机高的回答人可能愿意回答几个开放式问题，但是，如果研究者在设计问卷时以开放式问题【Open-ended Questions】为主，他就会发现很少有问卷能完成并返回，而那些返回的问卷仍可能有大量的缺失数据【Missing Data】或无关数据。第三，问卷必须独立使用。换言之，由于没有访问员帮助澄清填答说明或提供额外信息，在问卷上就必须给出可能的回答人用以回答问题所需的所有信息。第四，为了简化回答人的填答，调查者就不能使用分支问题或跳答。每个问题都应包含回答选项，而这些回答选项能让每位回答人可以轻松地用来描述自己的态度、行为、知识或特征。在一些场合下，这意味着在回答选项中必须包含一个“不适用(not applicable)”的选项。

无督导的自填式问卷的最大填答弱点在于，一旦问卷离开了研究者的办公室，研究者就失去了对问卷的控制，包括问卷到底由谁完成，以及在完成问卷时回答人是否咨

询了其他人。同样,研究者也无法控制问题的回答顺序。结果就是,当一组问题很可能对另一组问题产生混淆、偏差【Bias】或影响其回答时,就不能使用自填式问卷。

自填式问卷的设计

在决定数据是否可以通过邮寄或其他自填式问卷收集时,必须评估三种信息——目标群体的读写水平、目标群体的填答动机及研究问题【Research Question】的特征是否适合使用自填式问卷去收集数据。读写水平通常相当容易评估,填答动机则相当难以确定。要适应于研究,问卷的话题必须是那种可以在相当程度上由一份相对简短且集中的问卷承载的,并且它的问题要同样本中的每个人都有关。然而,自填式问卷特别不适用于以下情形:探索性研究,或者,当调查者正处在发展某个研究问题和确定所要研究问题的过程时。

对自填式问卷而言,尤其重要的是要做到使用者友好。而这既要通过问卷的写作方式,也要通过问卷的呈现样式,才能达到最佳程度。问题务必简短、具体,不能泛泛而言。这可能意味着必须问连锁问题,而不是单个问题。回答选项必须穷尽且互斥。有时,这意味着尽量以标准的方式制作问题,以便让回答人能选择多个答案来描述自己,也意味着必须在回答选项中列入一个"其他(residual other)"选项。当研究者认为可能列出的回答选项并没有穷尽时,其他选项尤为重要。

设计问卷的最大错误之一是试图让它们看上去简短,认为这样可以增加回答率、减少邮寄费用。这样的做法包括,所有的问题压缩到两页纸上、压缩或取消问题的间距,以及使用小号字。实际上,这些做法只会减少回答率。把问题压缩到很少的几页纸上一般意味着使用小号字和不给问题留间距。小号字会造成回答人阅读问卷的困难。间距的合理使用能让回答人更容易看清问题和相关的问题选项,且更方便回答人快速地完成整个问卷。

——Linda B. Bourque

(王玥译 赵锋校)

参考文献

Bourque, L. B., & Fielder, E. P. (2003). *How to conduct selfradministered and mail surveys* (2nd ed.). Thousand Oaks, CA: Sage.

Couper, M. P. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, *64*, 464-494.

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method*. New York: Wiley.

自填答问卷(Self-Completed Questionnaire)

参见自填式问卷【Self-Administered Questionnaire】。

自陈式测量(Self-Report Measure)

与个体有关的信息既可以通过观测【Observation】得到,也可以通过调查问卷由调查对象报告自己的行为或态度。在某些情况下,研究者可能要选择使用哪一种观测

方法，但是在许多情况下，自陈式测量是唯一能够使用的方法。研究者可以请个体站到秤上去称他们的体重，也可以在一份调查【Survey】问卷中问他们有多重。在这种情形下，秤更可能给出正确的值。对投票行为感兴趣的研究者则可能核对官方的记录，或在调查问卷中询问回答人在最近的选举中是否投票来获得投票信息。有些投票研究同时使用两种测量方法，以确保对回答人实际行为的效度【Validity】更有把握。

自陈式测量的潜在问题是它们可能包含一种或多种系统误差【Systematic Error】或者偏倚【Bias】。这可能源于问题的措辞方式、给予回答人的回答选项、所问问题的顺序、回答人与访问员【Interviewer】的互动，或者特定社会环境下回答人对“可接受(acceptable)”答案的看法。

回答人用自己的词汇作答的开放式问题【Open-Ended Questions】，通常比回答人必须从所给选项中回答一个封闭式问题【Closed-Ended Questions】会给出更多不同的答案。给予回答人一个明确的机会说“我不知道”，或选项中有“我还未真正想过”，相较于不给这类选项而言，常产生不同的回答模式。量尺的范围以及在不同问题间它们的差别也能解释回答人所给答案的差异(Schwarz, Hippler, Deutsch, & Strack, 1985)。

所问问题的顺序可能导致回答人在一系列问题中的最后一个题目给出一个无可避免的答案。询问一组同意或不同意式样的问题可能引致反应定势【Response Set】偏倚，因为回答人乐于同意呈现于他们面前的态度陈述(Schuman & Presser, 1996)。

一种最重要的自陈式偏倚与社会期望偏倚【Social Desirability Bias】有关，即相比于其他答案，回答人倾向于给出让人感觉更能为社会接受的回答。这是对一种普遍观察到的倾向，即人们说他们投票了，但记录核实表明他们中的许多人没有投票的最常见的解释。由于同种原因，种族主义态度通常会被低报。这些自陈的回答模式还关联于回答人的社会特征与访谈员的社会特征之间的关系性质，诸如两者间的人种关系和性别关系。

——Michael W. Traugott

(马妍译　赵锋校)

参考文献

Schuman, H., & Presser, S. (1996). *Questions & answers in attitude surveys*. Thousand Oaks, CA: Sage.

Schwarz, N., Hippler, H.-J., Deutsch, B., & Strack, F. (1985). Response scales: Effects of category range on reported behavior and comparative judgments. *Public Opinion Quarterly*, *49* (3), 388-395.

Traugott, M. W., & Katosh, J. P. (1979). Response validity in surveys of voting behavior. *Public Opinion Quarterly*, *43*(3), 359-377.

语义差异量表(Semantic Differential Scale)

语义差异(semantic differential, SD)是由奥斯古德在1940年代和1950年代发展的一项旨在定量的测量【Measure】语言意义的技术。对不同的个体，词汇作为他们在世界上的经历的函数可能有不同的含义。例如，“贫穷”对于7~70岁的人有不同的含义，对于富裕的人和无家可归的人也有不同的含义。他们对贫穷理解的表达受到他们

经历的调节。SD 通过赋予我们对词汇的不同理解以某些精确的内容来捕捉这些不同的意义。

典型的语义差异测试要求一个受试者依据一系列描述性的两级(如好与坏)量尺来估计一个刺激词。受试者被要求在两个极端和对立的形容词之间(它们界定了量尺的端点)给刺激评分。通常,这些两级量尺有 5 个或 7 个点。奇数的点允许受试者选择一个中点或者中立点(如果使用偶数选项,就不容许作这种选择)。各种量尺的位置可能是未标记的、数字化的或者标签化的,见表 1。

表 1

好	[ ]	[ ]	[ ]	[ ]	[ ]	坏
好	-2	-1	0	1	2	坏
好	非常好	相当好	两者都不是	相当好	非常好	坏

回答者似乎偏好标签化的样式。

通常,一个刺激将依据大量如下这类两级量尺而予以评估。例如:

好—坏	硬—软
粗糙—光滑	快—慢
新鲜—陈旧	嘈杂—安静
小—大	年轻—年老
强—弱	紧张—轻松
主动—被动	有价值—无价值
冷—热	不公平—公平
诚实—不诚实	空的—满的
苦—甜	残酷—仁慈

对这些两级量尺的响应模式确定了受试者赋予刺激词的意义。

奥斯古德对语义差异技术的实用性的早期研究指出,回答人在这些两级量尺上对刺激的评分趋向于具有相关性。奥斯古德论说,虽然语义空间是多维的【Multidimensionality】,但是可以通过有限数量的正交维度【Dimension】而予以代表。通常来说,三维正交即可说明出现在对一组个别的两级量尺的响应中大部分共变。

这三个维度被标记为评估、潜能和行动(evaluation, potency and activity, EPA),已被运用于许多不同的研究中。对这一发现的一种解释是,虽然两极量尺的意义有变化,但是许多量尺的意义是基本等价的,因此,可由一组数量少的维度予以代表。在不同的群体、空间或时间之间,这些意义多大程度上是等价的,即群体、文化和时代的相似性和差异性是可以检验的。

在实践中,语义差异可以用于发展对态度的具有辨识力的理解。例如,美国人经常对他们的国家机构和政治领导持有模糊的见解。然而,通常只用一个简单的问题,请市民对某一政治制度或某个民选官员作出评价。用来调查公众对美国总统认知的一个常见问题是"你是否同意或反对(现任总

统的名字)作为一个总统处理其工作的方式?"运用语义差异方法,通过要求被访者在一系列两极量尺中评价总统,可能发展出一个更复杂的舆论评定,见表 2。

表 2 请在以下每一个方面评价乔治·布什

好	[ ]	[ ]	[ ]	[ ]	[ ]	坏
弱	[ ]	[ ]	[ ]	[ ]	[ ]	强
主动	[ ]	[ ]	[ ]	[ ]	[ ]	被动
公平	[ ]	[ ]	[ ]	[ ]	[ ]	不公平
有益的	[ ]	[ ]	[ ]	[ ]	[ ]	无益的
不诚实	[ ]	[ ]	[ ]	[ ]	[ ]	诚实
尖锐	[ ]	[ ]	[ ]	[ ]	[ ]	平淡
无效率	[ ]	[ ]	[ ]	[ ]	[ ]	有效率
快	[ ]	[ ]	[ ]	[ ]	[ ]	慢
有价值	[ ]	[ ]	[ ]	[ ]	[ ]	无价值
无权力	[ ]	[ ]	[ ]	[ ]	[ ]	有权力
勇敢	[ ]	[ ]	[ ]	[ ]	[ ]	懦弱

对每个量尺的回答既可以个别地予以检验,也可以运用数据缩减技术(如因子分析【Factor Analysis】),在 EPA 维度上来探测公众对总统的评价。

语义差异在社会学、沟通研究、认知心理学和市场研究中都有广泛应用。

批评者发现了语义研究方法在测量态度时的一个逻辑陷阱:基于语义差异研究的最初假定,对一个刺激项(即一个词),人们持有不同的态度/知觉。但是 SD 技术又假定,被选中用以代表两级量尺的端点的形容词对每位评分者具有相同的意义。

——John P. McIver

(马妍译 赵锋校)

参考文献

Heise, D. (1971). The semantic differential and attitude research. In G. F. Summers (Ed.), *Attitude measurement* (pp. 235-253). Chicago: Rand McNally.

Osgood, C. E. (1952). The nature of and measurement of meaning. *Psychological Bulletin*, *49*, 197-237.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.

Osgood, C. E., & Tzeng, O. C. S. (Eds.). (1990). *Language, meaning, and culture: The selected papers of C. E. Osgood*. New York: Praeger.

Snider, J. G., & Osgood, C. E. (Eds.). (1969). *Semantic differential technique: A sourcebook*. Chicago: Aldine.

半对数的(Semilogarithmic)

参见对数【Logarithm】。

符号学(Semiotics)

宽泛地界定,符号学是关于意义的理论,即意义和交流(communication)系统的结构和功能的科学领域,以及由它们所构建起的含意(有意义的)整体的理论。符号学(semiotics)一词源自 semiotic(符号的),由现代符号学的两个创立者之一,美国哲学家查尔斯·皮尔士(Charles Sanders Peirce, 1839—1914)首先使用。另一个竞争词是符

号科学(semiology),由符号学的另一个创立者瑞士的语言学家费尔迪南·德·索绪尔(Ferdinand de Saussure,1857—1913)发明。1964年在伯明顿的印第安纳大学召开了一个会议之后,符号学为人所普遍使用(尽管索绪尔派传统的一些符号学家继续在两个词之间做区分)。

根据这一定义,符号学不是一个独立的科学,而是一个科学的领域。实际上,符号学的目标,即含意整体与系统,涉及面太广,而不能界定为一门科学;确实,它跨越了众多已有的科学(如语言学、人类学)和学科(如文学、建筑学、音乐)。

就索绪尔派结构语言学【Structural Linguistics】、皮尔士式的符号学,以及其他大部分的符号学理论而言,符号学的基本单位是单位示号(sign)(尽管有一些理论针对更广的含意实体)。单位示号是任何只要能被用作一个物以表现(意指)另一物的事物。“树(tree)”这个词,就是一个单位示号,指向一些确定类别的植物(有树干、树枝、树叶等)。因此,单位示号是含意单位,由一项能指(signifier)(索绪尔的术语)——意义的媒介——和一项所指(signified)——意义本身所构成。符号学研究两个层面(至少)间的关系——能指和所指,而这一事实使符号学有别于信息理论,因为后者仅涉及能指层面。皮尔士根据一个三元结构关系,对示号做了区分,包括:示号,即任何传递意义的事物;对象,示号予以说明的事物;以及解释项,由对象引起的想法。解释项本身可以由其他示号命名,引出它自身的解释项,如此这般(理论上)形成一个无限进行的“无限示意(unlimited semiosis)”过程。对皮尔士而言,示号的最基本区分有图示、指示、意示。其中指示和图示与他们意指的东西实际上相关或存在相似性,有别的是,意示仅仅是约定俗成的。在这一点上,皮尔士的意示类似于索绪尔约定的和任意的符号,此即,在符号和它们所代表的东西之间没有天然的关联,同样,这一原则适用于能指和所指间的关系。

在交流过程中,示号被合入含意整体(signifying wholes),诸如信息、言语或文本【Texts】。一个含意整体是根据确定的句法规则(syntactic rules)组合起来的一群示号。一座建筑就是一个含意总体的例子;某一确定类型的建筑(带尖顶、塔和一系列其他特征)标志着“哥特式教堂(gothic cathedral)”,而由后者可能令人想起“神圣之地(sacred place)”和“崇高(transcendence)”等含义。这一例子同样显示了一个示号的两层意义:字面的意义或显意(denotation),以及间接的或象征的意义,隐意(connotation)。

含意整体预设了意义系统的存在。索绪尔称意义系统为语言(langue)[通常译作“语言系统(the language system)”,用在符号学中,指所有示号的系统]。该系统是共时的(synchronic),此即,它代表了语言的一个特定的静止状态[与它的历时的(diachronic)变化相反];它本质上是社会性的,对索绪尔而言,相对于自然语言的研究,它是首先被认可的观点。结构(structure)一词经常被用作系统(system)的对等物,尽管它也用来指统摄系统的规则。

系统是索绪尔符号学的关键概念。若依照示号的任意性原则,意义则不源于任何示号同真实世界之间的关系,而是在某个抽象的、纯粹形式的、具有代数性质的系统之内,示号之间的相互关系。能指(和所指)通过它们与系统的其他能指(或所指)之间的差异而构成。意义的区别性质是索绪尔派结构语言学的基础原则之一。最后,否认所指有确切的内容,因为意义随示号同系统中其他示号的关系而变化其值(value)。值——以及整个系统——是通过指者间两种关系的互动而确定的:词意(paradigmatic)关系,源自示号的分类集(索绪尔的一个集的例子是

"教学""指导""学徒""教育"),以及句意(syntagmatic)关系,从各种词形表中选出示号,再将它们组合成线性的语言结构单位(syntagms)(如句、篇章)。

语言(langue)是与言语(说话)相对的,而后者是由个人对语言系统的使用。语义系统在社会使用,并在社会中发挥功能。它们的核心功能在于交流,尽管这并不是唯一功能。含意整体建立在言语的层面上。语用学(广义上)是符号学的分支,旨在研究含意整体的生产过程和作用过程,即它们的使用;它们对于接受者的认知、情感或行为—身体的影响;以及它们与交流情境的动态互动。

历史发展

尽管古希腊和其他古文化关注过意义的问题,现代符号学源自皮尔士和索绪尔的工作。在美国,符号学通过查尔斯·威廉·莫里斯的工作获得了广泛的认可。他把符号分析分解为句法学、符号学和语用学——他将其定义为关于示号和对话者之间关系的研究。美国的一个主要的趋势是把符号学从对人类意指系统的研究扩展到动物交流(动物符号学),以及进一步扩展到所有生物有机体的自然过程(生物符号学),而这些在很大程度上要归功于托马斯·谢伯克的工作。

俄罗斯语言学家罗曼·雅格布森对欧洲符号学的形成有决定性的影响。1920年代,在索绪尔派传统下工作的雅格布森,参与了俄罗斯的两个符号学团体的创立。这两个团体形成了俄罗斯形式主义的核心。1930年代,他推进了布拉格语言学圈的建立,而这个学圈恰标志了欧洲语言学的结构主义【Structuralism】的成立。雅格布森和布拉格学圈的工作在于,将自然语言设想为多种示号系统之一,而许多示号系统共同构成了文化,形成了一个复杂的交流系统(这一观点后来在第二次世界大战后由Moscow-Tartu符号学派继承)。丹麦语言学家路易斯·耶尔姆斯勒夫也对欧洲符号学的发展作出重要的贡献,他在1931年与他人共同建立了哥本哈根语言学圈。

第二次世界大战期间,雅格布森遇到并影响了一位法国的人类学家,克劳德·列维-斯特劳斯。克劳德·列维-斯特劳斯的结构人类学对符号学在1960年代和1970年代的兴盛至关重要。1964年,文学学者罗兰·巴尔特发表了第一篇当代的符号学论文,《符号学基础》(*Eléments de Sémiologie*),很快其他知名的符号学家也发表他们的主要著作,例如,巴黎学派的创始人阿尔吉尔达斯·格雷马斯和意大利哲学家翁贝托·艾柯。与此同时,这一趋势在巴黎继续发展,即后来著名的后结构主义【Poststructuralism】;其中杰出的作者有哲学家雅克·德里达和弗朗西斯·利奥塔,心理分析家雅克·拉康,知识["言语(discourses)"]历史学家米歇尔·福柯,以及社会学家让·鲍德里亚。

方法论和应用

与两种现在最常使用的意义分析的方法——内容分析【Content Analysis】和语义差异【Semantic Differential】——相比较,则可以理解符号学方法。两种方法都预设了定性类别的创造,并且其结果都体现于定量的统计分析;然而,它们采纳了相当有限的文本观,也缺少任何实质性的理论营造,从而能够从它们的方法论成果中获得其价值,更不必说技术了。与此相反,符号学方法以非常成熟的理论为基础,涉及一个文本的所有或几乎所有面向,已被确立为首要的定性方法。

意义分析的传统方法一般不处理结构

段的或线性的文本维度。另一方面,符号学容纳了一个相当完备的主要由格雷马斯构想的叙事(narrative)理论。格雷马斯将一个文本分解为各种"做(doing)"的单位。这些"做"的单位又通过联系或分离多种主体和客体,从而形成诸"状态(state)"的单位。这一基础的叙事程序,动作(the act)产生或改变一种状态,动作影响了一种状态到另一种状态的过渡。这些动作的结构段组织(syntagmatic organization)带来了行动(action)。每一个言语——不仅是叙事言语——都是一个从初始状态走向最终状态的过程,并且可以通过变换的算法【Algorithm】予以形式地模拟。

巴尔特(Barthes,1977)给出了符号学方法的一个最早的实例,即对意大利食品生产商 Panzani 的广告的语用学分析。语用学分析的目标类似于内容分析和语义微分的目标。巴尔特区分了广告中两种互补的信息(messages)类型,即图示的和语言的。在图示的信息中,他识别出两个层次,字面的或外延的信息(例如图像中的对象),以及象征的或内涵的信息;语言的信息服务于引导对象征的解释。

巴尔特观察到作为结构段的字面信息是象征信息的必要支撑,因为前者的某些不连续性的元素引起"强的(strong)"示号的出现,而"强"的示号是第二层面意义的载体。他识别出此一类型的四种主要意义:半开的网兜中载满了溢出来的新采购的产品和家庭用品,兴高采烈地"从市场回家";"italicity"(Panzani 的缩写名字),一种所有意大利事物的精髓——从意大利面到画,由三种颜色绘出,并伴以(地中海的)蔬菜、番茄和青椒(绿色、白色和红色是意大利国旗的颜色);由于汇集了许多不同的食物,含意是"完全的烹饪服务";以及由整个组合,即画面的怀旧情调显示的"宁静生活(still life)"。这四种象征意义对应于四种语用学准则(codes),即巴尔特所说的实践的、民族的、文化的和审美的。

尽管符号学分析通常是定性的,然而偶尔也可以同定量的分析结合在一起。在研究希腊北部四个地区的区域空间(regional space)概念时,亚历山德罗斯·拉戈普罗斯和卡林·拉戈普罗将他们访谈内容析分为 32 个语用学准则,再组成 8 个主要的集,即"有关空间的言语"的特征。他们使用频率计数和列联表来研究一组广泛地含有准则的语义(概念的)变量,以及这些变量同社会学的变量(社会阶级、性别、年龄、教育)和社会地理学的变量(城乡、社区等级)的关系,以及各种语义数据间的关系。

今日符号学

如今,符号学已经遍布所有社会和人类科学,是我们这个时代的支配范式——后现代主义【Postmodernism】——的重要组成。在语言学和文学研究中,它助力了词汇学、话语分析【Discourse Analysis】(特别是修辞学)、叙事学,以及文体学的发展。它还深刻地影响了人类学和神话学的研究。它同样也渗透到符号学中,为非言语和混合系统的研究作出重要贡献,引出了各种符号学的创造,例如行为、绘画与摄影、音乐、建筑学、城市空间、烹饪习惯、竞技活动、时尚、广告、电影、戏剧等诸多符号学。最重要的是,符号学为文化现象的比较研究提供了一个统一的理论框架,极大地促进了当今无比繁荣的含意结构的多学科和跨学科研究,以及各种实践的兴盛。

——Karin Boklund-Lagopoulos
Alexandros Lagopoulos
(王玥译　赵锋校)

参考文献

Barthes, R. (1967). *Elements of semiology*. London:

Jonathan Cape.

Barthes, R. (1977). Rhetoric of the image. In *Image, music, text* (S. Heath, Trans.). London: Fontana.

de Saussure, F. (1979). *Cours de linguistique générale* (T. de Mauro, Ed.). Paris: Payot.

Eco, U. (1976). *A theory of semiotics*. Bloomington: Indiana University Press.

Greimas, A. J., & Courtés, J. (1982). *Semiotics and language: An analytical dictionary*. Bloomington: Indiana University Press.

Lagopoulos, A. Ph., & Boklund-Lagopoulos, K. (1992). *Meaning and geography: The social conception of the region in northern Greece*. Berlin: Mouton de Gruyter.

Peirce, C. S. (1931-1958). *Collected papers of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press.

半偏相关(Semipartial Correlation)

半偏相关是一个统计量,其作用在于衡量控制分析中其他预测变量的效应,一个因变量【Dependent Variable】(或准则)同一个调整后的自变量【Independent Variable】(或预测变量)的关系。类似于偏相关【Partial Correlation】,在多元回归分析【Multiple Regression Analysis】的过程中,需要计算半偏相关。

图1说明了半偏相关同偏相关和皮尔逊相关在概念上的区别。图中灰色的圆圈(A和B)代表预测变量,而黑色的圆圈(Y)代表效标变量。盒形代表减去了B对A和Y的影响后,A和Y的残差。曲线箭头代表图中每对变量之间的相关关系(或协方差【Covariance】),而直线箭头代表一条连接B到Y的回归路径。

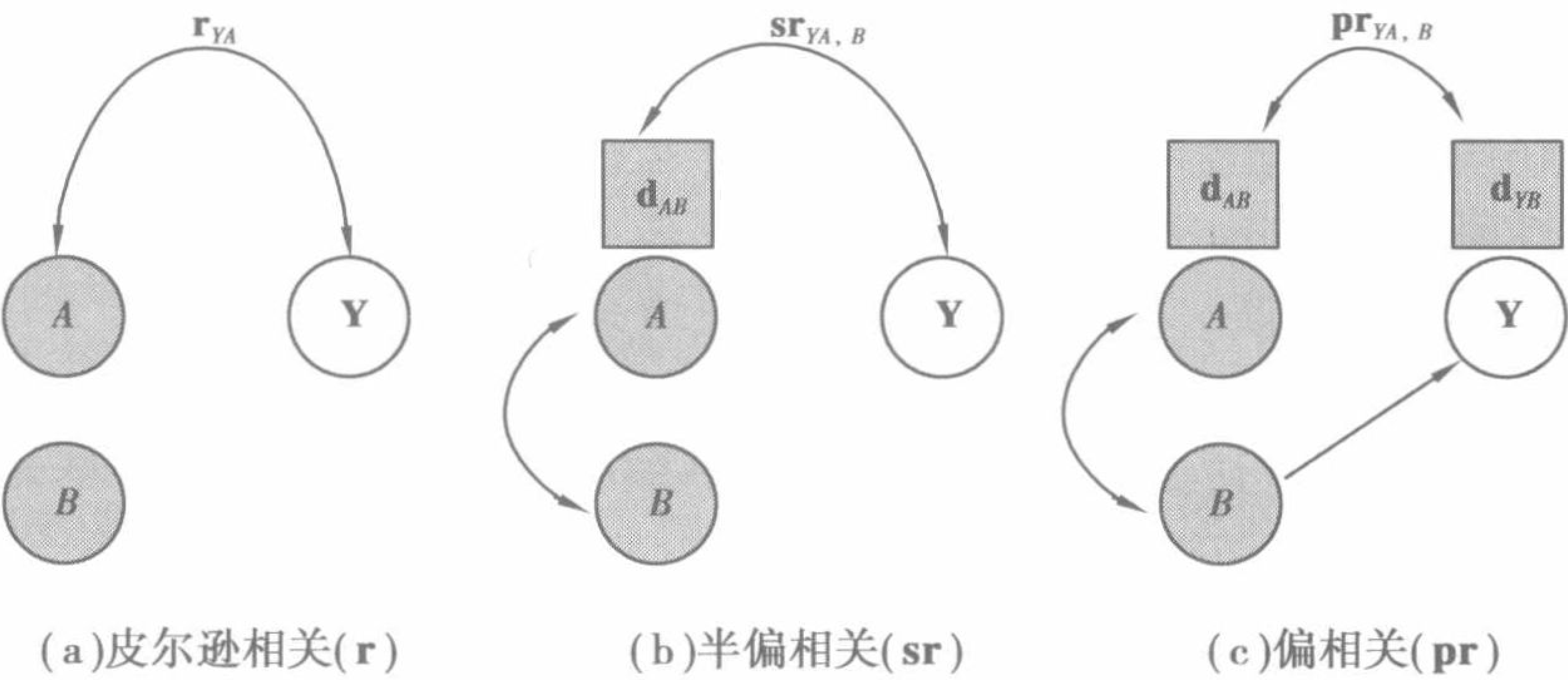


图1 偏相关和皮尔逊相关之间概念差异的说明图

在图1(a)中,曲线箭头$\mathbf{r}_{YA}$代表一个零阶的皮尔逊相关系数,而它在不考虑第三个变量B的效应的前提下,衡量了A和$\mathbf{Y}$之间相关的程度。在图1(b)中,曲线箭头$\mathbf{sr}_{YA,B}$表示一个半偏相关系数。它衡量了效标变量$\mathbf{Y}$和A的残差【Residual】(即$d_{A,B}$)(通过从A的方差中减去A与B的协方差得到)彼此的相关程度。在图1(c)中,曲线箭头$\mathbf{pr}_{YA,B}$代表一个偏相关系数。它衡量了$\mathbf{Y}$的残差(即$d_{Y,B}$)与A的残差(即$d_{A,B}$)彼此相关的程度。

很清楚,如果在一个多元回归分析中,$\mathbf{Y}$是通过预测变量A和B进行预测的效标变量,那么,从A和B得到的半偏相关和偏

相关都会同 A 和 B 对 $\mathbf{Y}$ 的具体(或唯一)贡献分别地联系在一起。特别是,A 的半偏相关的平方是 $\mathbf{Y}$ 的总方差中唯一的由 A 解释掉的比率,即 A 从中除去 A 与 B 的共享方差后,A 中残留的方差同 $\mathbf{Y}$ 的总方差之比。不同地,A 的偏相关系数的平方是 $\mathbf{Y}$ 的残差唯一的由 A 解释掉的比率,即同时从 A 和 $\mathbf{Y}$ 中除去 B 的效应后的比值。

对任意给定的两个变量之间的偏相关和半偏相关系数,都可以根据下列算术式,用这对变量的零阶皮尔逊相关系数进行计算:

$$\mathbf{sr}_{YA,B} = \frac{\mathbf{r}_{AY} - \mathbf{r}_{AB} \times \mathbf{r}_{YB}}{\sqrt{1 - \mathbf{r}_{AB}^2}} \quad (1)$$

$$\mathbf{pr}_{YA,B} = \frac{\mathbf{r}_{AY} - \mathbf{r}_{AB} \times \mathbf{r}_{YB}}{\sqrt{1 - \mathbf{r}_{YB}^2} \times \sqrt{1 - \mathbf{r}_{AB}^2}} \quad (2)$$

比较等式(1)和等式(2),半偏相关系数(如 $\mathbf{sr}_{YA,B}$)的计算在分母的构成上不同于偏相关系数(如 $\mathbf{pr}_{YA,B}$)的计算。偏相关等式的分母项要小于半偏相关等式的分母项(亦即,就偏相关而言,其分母项是两个小于1的项的乘积,而半偏相关的分母项只包括一个小于1的项)。两者的差异使得半偏相关系数总是小于偏相关系数,除非 A 和 B 是相互独立或者不相关的。

——Marco Lauriola

(马妍译　赵锋校)

参考文献

Darlington, R. B. (1990). *Regression and linear models*. New York: McGraw-Hill.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd ed.). New York: Holt, Rinehart and Winston.

半结构化访谈(Semistructured Interview)

半结构化访谈是一个涵盖广泛的术语,常用于描述与定性研究【Qualitative Research】相关的一系列不同形式的访谈。不同于结构式访谈【Structured Interview】——其中包含了一个以同样的方式向所有访谈对象提出问题的固定次序,半结构化访谈的确定特征在于它们具有一个弹性的和可变动的结构。半结构化访谈的结构常依照一个辅助的备忘录或访谈提纲【Interview Guide】而加以组织。这一结构涉及贯穿访谈全过程的话题、主题或各个部分,而不是一个依次排列的标准化问题的问卷。其目的常在于要确保访谈的弹性,即根据具体情况确定提问的方式和提问的顺序,根据不同的访谈对象确定某一部分问题是否适用于他们,是否需要他们进一步作答。这就是为什么访谈对象自身的理解可以同研究者的兴趣一道塑造访谈,也是非预期的主题得以突现的缘由。

半结构化访谈的逻辑和基础

半结构化访谈与社会科学领域的诠释主义【Interpretivism】、互动论【Symbolic Interactionism】、建构主义【Constructionism】、女性主义【Feminism】、心理分析【Psychoanalytic】、口述【Oral】或生活史【Life History】的传统紧密相连。它反映出同人的知识、理解、解释、经验及互动相关的一种本体论【Ontology】立场。在当前的半结构化访谈实践中,人们可以辨识出不同传统的要素。而这些实践,根据珍妮弗·梅森的建

议,具有以下广泛的特征:互动式的对话;相对非正式的风格;主题性的、以话题为中心的、传记式的或叙事的实施方式;以及相信知识有其情境和背景,因此访谈的功能在于确保能将相关的背景置入研究的关注点,并由此生成情境化的知识(Mason, 2002, p. 62)。

大多数评论者同意,半结构化访谈的逻辑是要以互动的方式生成研究资料,而斯丹纳·苛费尔(Steinar Kvale)还把定性研究访谈描述为“一个知识的构建场所”(Kvale, 1996, p. 2)。这意味着在知识构建过程中,不仅是访谈对象,访谈者也被视作一个积极的、反思性的(reflexive)角色。于是,资料就被视为源自访谈者和访谈对象间的互动,而研究者要去分析的是这种互动,而不仅仅是访谈对象给予的“答案(answers)”。

这种以相对开放、弹性和互动方式形成的访谈结构的通常目的是要生成访谈对象对他们自己的视角、感受、经验、理解、解释及互动的说明。更加标准化的和结构化的方式可能过度地把研究者自己的意义和理解框架强加在之后的资料中。此外,它还可能忽略由访谈对象的观点看具有重要性的事件和经验,而它们虽然同研究相关,却不在研究预期中,或是忽略那些对访谈对象自己的传记或他们知觉事物的方式具有特殊意义的事件和经验。

有时,半结构化访谈还暗含一种对研究伦理【Ethics】的特殊处理,即尽可能平等地安排访谈者和访谈对象间的权力关系,也就是让访谈对象有充分的机会用自己的方式讲自己的故事。尽管这可能让半结构化访谈显得比标准的或结构的方法更加符合伦理,但也必须承认,生成的资料可能是非常私人的和敏感的,特别是当它们来自访谈者与访谈对象间的信任和亲密融洽【Rapport】的关系时。因此,在这个背景下,保密和伦理问题并不简单。

半结构化访谈的类型

实践上,存在着相当不同类型的半结构化访谈。不同类型间的差异在于:在何种程度上,访谈者负责组织和指导访谈互动;访谈的组织逻辑的性质,比如是一个生命故事、自传,还是一个或一组松散的话题;在何种程度上,可以将后续资料视作互动的结果或访谈对象的说明而加以分析;以及在何种程度上,访谈资料可以为自身讲话,或需要由谁及经由何种机制而予以解释。下面是三种不同类型的半结构化访谈:

1.*民族志访谈*通常是涵盖更广的田野研究的一个组成部分。民族志访谈的实施经常伴随着田野中不断发展的关系,因此各种形式的资料既有得自访谈之内的,也有得自访谈之外的。这种访谈能将访谈者与访谈对象之间一段时期内的持续联系作为访谈的基础,并因此可以增加知识的互动合作生产以及融洽双方关系的可能性。在分析后续资料时,它也非常强调访谈对象自己的解释和理解。

2.*心理分析访谈*有时被当作一项治疗活动的组成部分,尽管这不是访谈的必要构成。它的目的通常既要去探寻访谈对象的意识又要探寻他们的潜意识理解。温蒂·霍尔韦和托尼·杰斐逊(Hollway & Jefferson, 2000)建议使用自由联想叙事法(the free association narrative method),即使用一种最低限度的访谈者—施加的结构(interviewer-imposed structure),并通过分析访谈对象自由联想的模式,从而促进研究者对潜意识的主体间性和心理社会主体的理解。因此,研究者并不将访谈对象自己的解释当作事实的解释,而是动用心理分析理论去分析资料。

3.*生活史或口述史访谈*包括编纂自传和生活故事。研究者可以用多种方式编纂生活史,包括使用文献法或可视方法,如对日记、信件及照片的分析。它们都可以构成访谈的一部分,例如,请访谈对象就家庭的照片聊一聊。其中,有些生活史以一种完全不加解释的方式,即依照访谈对象的讲述,使用相关方法来记录他们的生活。另一些则强调生活史知识的构建过程中解释和互动的成分。像民族志访谈一样,生活史访谈有时也随着一段时期内不断发展的关系而同时开展。

对半结构化访谈的评价

已经证明半结构化访谈是一种有价值的获取阐释性资料的方法。有些批评认为,由于它所生成的资料是非标准化的,所以这些资料不容许进行访谈间的比较,而这些批评是错误的。因为访谈资料比较的逻辑要求将比较建立在对每个个案全面理解的基础上,而不是基于所有案例数据的标准化。然而,单独使用半结构化访谈可能生成的只是片面的阐释性理解,而其他方法能够成为有益的补充,如那些能拓展知识的情境维度的方法,包括参与观察【Participant Observation】和可视方法【Visual Research】。

——Jennifer Mason

(王玥译　赵锋校)

参考文献

Hollway, W., & Jefferson, T. (2000). *Doing qualitative research differently*. London: Sage.

Kvale, S. (1996). *InterViews: An introduction to qualitative research interviewing*. London: Sage.

Mason, J. (2002). *Qualitative researching* (2nd ed.). London: Sage.

Plummer, K. (2001). *Documents of life 2: An invitation to a critical humanism*. London: Sage.

研究敏感话题(Sensitive Topics, Researching)

研究者常常必须以某种方式研究那些敏感话题。然而,并不存在一个敏感话题的确定清单。通常,敏感性似乎经常产生于要研究的话题与所研究的话题社会背景之间的关系。有些看似困难的话题可以无问题地被研究,而有些表面上无害的话题却可能是高度敏感的。因此,那些起初看上去陌生、古怪甚至是令人恐怖的群体,有时会感到被广大社会所误解,同时,会平静地看待一位研究者的出现,并认为这位研究者可以将他们的世界观传递给更广大的听众。另一方面,一些表面上正常的环境可能隐藏了那些参与者希望规避审视的越轨模式。在这些情况下,调查通道【Access】和田野关系【Field Relations】都可能成为预期之外的问题。

敏感话题经常包含所研究对象面临一定程度的威胁或风险,即信息提供者质疑研究数据的收集、保存,以及(或是)发布。这些威胁经常产生于研究者对社会生活特定领域的兴趣。当研究触及生活的如下方面时,它可能是敏感的,如性活动、失去亲人,或被强烈保有的、私人的、折磨人的或神圣的个人信念。被研究者眼中的如下可能性也是敏感性的一个常见来源,即研究可能暴露研究参与者的相关信息,而这些信息会以某种方式导致污名或加罪。当研究撞上了有权力的人或机构的既得利益,或者压制或统治的活动时,研究也常是敏感的。研究也可能对研究者有威胁。那些研究越轨群体的研究者,有时要共享那些贴在被研究者身

上的社会污名。朋友、同事,以及大学的行政人员有时认为,调查性行为问题的研究者也拥有他们的研究对象的偏好。在暴烈的社会冲突环境中,或者当要研究的群体处于暴力经常发生的环境中,即可能出现身体的危险。

因为它们通常涉及那些其他研究中不常见的意外情况,所以敏感话题既提出了困难的方法论问题,也提出了技术问题。进入田野是常见的问题。那些控制进入田野的人经常拒绝给研究者提供帮助,或设置进入田野的条件或约定。在敏感的背景下,研究者与被研究者之间的关系可能以种种方式受到不信任、隐瞒和掩饰的阻挠,而这又会影响数据的信度和效度。在这些情况下,对研究者的接纳通常要依赖于同研究参与者的信任关系的长期培育。就这些由研究过程的威胁特征而导致的问题,找到技术上或方法上的解是可能的,如在极强的匿名条件下收集数据。尽管会涉及严肃的伦理问题,一些研究者还是采用了某些强健的调查策略,其中包括为了克服有权力的人群隐瞒他们的活动,不被外界审视,使用隐蔽调查法或说服式访谈策略。

敏感话题有时会出现与研究伦理、政治和法律方面相关的难题。那些研究敏感话题的人可能需要比研究那些一般无害话题的人,更加敏锐地意识到他们对研究参与者的伦理责任。需要特别留意,例如,匿名化和数据保护。研究参与者已经逐渐加强了针对研究者的法律权利。许多国家现在有法律体系,有权对社会研究予以事先的伦理审查。不过,有时,法律框架阻碍研究敏感话题的方式,恰同制定这一框架的精神相背离。目前,研究者常常进入合作式的研究关系中,而这些关系容许无权力的或处于边缘的人群就他们的担忧和渴望发出自己的声音。参与行动研究【Participatory Action Research】方法容许发展研究者同研究参与者之间平等的和互惠的关系。不过,这种关系并不能免于研究者和被研究者之间的紧张,或是免于面对共同选择的压力。

敏感话题考验研究者的才智,还可能对他们作出实质性的个人要求。技术、毅力和想象力都是成功解决敏感话题引发的问题所需要的。尽管这些已经鼓励研究者在某些领域中的创新,如在问卷调查中询问敏感问题、保护敏感数据,不过,研究中方法或技术的决策同伦理的、法律的或政治的考量之间的关系通常是复杂的。尽管有些研究者在研究敏感话题时,通过整个地退出该话题领域来应对研究中的难点,然而,敏感话题始终是社会研究中既富于挑战性又具有必要性的活动。

——Raymond M. Lee

(王玥译　赵锋校)

*也可参见行动研究【Action Research】、匿名【Anonymity】、隐蔽研究【Covert Research】、研究中的危险【Danger in Research】、田野关系【Field Relations】、研究的政治【Politics of Research】。

参考文献

Lee, R. M. (1993). *Doing research on sensitive topics*. London: Sage.

敏化概念(Sensitizing Concept)

美国社会学家赫伯特·布卢默是符号互动论【Symbolic Interactionism】的创立者,他将敏化概念同一些定量社会研究者所坚持的确定概念对立起来。敏化概念不涉及

用“固定的和具体的程序”去辨别一组现象，而代之以给予“用以接近经验事实的一个一般参照和指导的感觉”。所以，确定概念“提供了看什么的处方，而敏化概念只是建议随着它去看的方向”（Blumer, 1969, p.148）。

布卢默承认理论概念的模糊性可能是个问题。然而，他否认问题的解决方法在于去发展确定概念，正如那些在1930年代、1940年代和1950年代曾经影响了美国社会学（参见操作论【Operationism】）的实证主义【Positivism】科学哲学所建议的那样。他所为之奋斗的敏化概念反映的是实用主义【Pragmatism】哲学，而他自己对方法论的思考正源于此种哲学。著名的实用主义者杜威主张，概念是以使用为目的的工具，同时，应根据它们的工具价值——是否能促使人们去理解世界，是否能让人们对世界采取行动——而予以判断。就布卢默的论证而言，同样关键的是如下观念：概念，如果要去抓住同它们相关的现象的性质，就必须在研究的过程中去发展、去界定。他提出，通过精确地表述概念的内涵来开始研究，就像许多实证主义者所提倡的一样，等同于一种虚假的科学性，同时，这种做法很可能会阻碍获取社会世界的真正知识这一任务的实现。与此同时，布卢默著作中以更加激进的观点建议：在某种意义上，模糊性可能是社会科学概念的一个无可避免的特征。一个原因在于，把握社会现象的多样性、过程性和偶然性的困难。布卢默提出，这似乎意味着不可避免地依赖于默许的文化知识，即依赖于理解【Verstehen】。因此他写道，只有物理对象可以“被转变到一个时空框架内，或被带入乔治・米德所说的触—视领域（touch-sight field）”，然而，社会现象的观察需要“观察者对行为情境中的关系有所感知，并以此为基础，运用观察者经验中的社会规范来做判断”。（Blumer, 1969, p. 178）。

布卢默承认，社会学的多数核心概念都满足一个重要的敏化角色。然而他坚持，只要有可能，研究者还要通过丰富他们在社会世界中相关领域的一手经验，来澄清这些概念。首先要做的是，通过他所说的自然主义的研究（定性的或民族志质询），在这些研究中，经验实例在它们的自然背景中，受到“一双能够分离（概念的）类属性质的眼睛”的“细致而又有弹性的审视”（Blumer, 1969, p. 44）。

——Martyn Hammersley

（王玥译　赵锋校）

参考文献

Blumer, H. (1969). *Symbolic interactionism*. Englewood Cliffs, NJ: Prentice Hall.

Hammersley, M. (1989). The problem of the concept: Herbert Blumer on the relationship between concepts and data. *Journal of Contemporary Ethnography*, *18*, 133-160.

语句完成测验（Sentence Completion Test）

语句完成测验需要填写语句中缺失的一个或多个词语。下面是一个简单的例子：

她给他倒了一杯咖啡并给他一片__________。

选择出语义上恰当的完形词汇的能力

被视为语言能力水平的体现。另一种测验样式要求给出两个缺失的单词或用一段带有多个缺失单词的段落做测验，以便测试受试者对因果关系或逻辑关系的理解水平。

语句完成或"完形填空"测验在教育测验中有广泛使用(例如，作为美国 SAT 测验，Graduate Record Exams 和 the Test of English as a Foreign Language 的一部分)。测验中，给学生一组可能的完形词汇。所有的词汇在语法上都是合理的，但是只有一个在语义上是恰当的。完形作为语言能力的一种测量的效度依赖于此一假设，即在自然语言中理解一个词语的意义需要许多其他语义相关词语的支持结构。如果上述假设未能满足，受试者就不能参加考试，因为可信的完形表现依赖于事先已经安排好的整体结构的程度，否则测验就没有可信性了。

在心理学研究中，很少给受试者提供多种完成语句，而是让他们按照他们认为恰当的方式完成句子。研究所关注的因变量【Dependent Variable】是他们对词语的选择，或者是他们完成语句的时间。

心理学的应用分为两种类型，其分类依据是:完形是被用作非语言心理状态的间接测量(如人格的一个面向)，还是作为研究语言过程本身的考察工具(如测验语句理解的理论)。在非语言的研究中，只有同先前收集到的完形规范(已经证明对测量受试者的心理状态是有效的)相比较，完形才提供了关于受试者心理状态的信息。

在心理语言学中，语句完形常与自由联想和触发任务配合使用，以期量化那些平时难以获得和无意识的语言预期(Garman，1990)。在一项联想任务中，受试者被给予一个单一的词语提示，再请他们说出脑海中所有出现的词语。在一个语言中，每一个词在一个提示词后出现在首批反应词中的概率可以通过一个很大的文本估计出来。而受试者倾向于给出那些在给定提示词条件下具有很高条件概率的词汇。有时，研究者会在测试词前，给受试者一个提示词。这个提示词可能与测试词有语义关联，也可能同测试词无语义关联。触发指同用无语义关联的词作提示词相比，用有语义关联的词作提示词带来的对测试词认识所费时间的减少。触发效应可以通过要认识的词汇同触发词汇之间的替代性水平来很好地预测。替代性是一个词语在不造成语句语义上的不协调的条件下被另一个词替代的容易程度的一个量度，此外，通过测量词汇的分布的类似性，可以将词汇的替代性予以量化(Redington & Chater，1997)。替代性说明，意义具有内在的关系，同时，它能形式化地检验如下假定，即对意义的理解预先要求实质的、周边的语言结构的存在(Cruse，1986)。

生理上，在触发任务和不协调替换语句上下文中的不相关的词语会触发 400 组大脑事件相关电位，拉长注视的时间，增加反应的时间。

因为联想任务是语句完形的压缩形式，触发结果是通过与语句完形紧密相关的替代性测量来预测的，语句完形很可能根本地依赖于语言背景的统计测量。从这个角度看，第二种语言语句完形任务的应用可测量受试者内化替代性直觉的程度，而替代性直觉是由潜在的第二种语言的统计学特征产生，同时，它还是学习第二种语言必需的副产品。

——Will Lowe

(王玥译　赵锋校)

参考文献

Cruse, D. A. (1986). *Lexical semantics*. Cambridge, UK: Cambridge University Press.

Garman, M. (1990). *Psycholinguistics*. Cambridge, UK: Cambridge University Press.

Redington, M., & Chater, N. (1997). Probabilistic and distributional approaches to language acquisition. *Trends in the Cognitive Sciences*, *1*(7), 273-281.

序贯分析(Sequential Analysis)

在社会科学研究背景下,序贯分析是指组织和分析行为数据的方式,其目的在于揭示观测到的行为中存在的序次模式或规律性。因此,不像对数线性【Log-Linear】或多元回归分析【Multiple Regression Analysis】,序贯分析不是一项特定的统计技术,而是既有统计技术的特定应用。

序贯分析的数据通常来自系统观测【Systematic Observation】(Bakeman & Gottman, 1997),其性质是定类的,并且可能数量巨大。用于分析序次数据的许多统计分析技术包括那些通常适用于定类数据分析【Categorical Data Analysis】的技术,如卡方或对数线性分析,不过,就序次分析而言,通常一项有效的策略是先从数据中得出序次模式的指标,再用适当的统计方法予以分析(Bakeman & Quera, 1995)。因此,一个序贯分析可以就观测到的行为的顺序提供基本的描述信息。与此同时,它还可以将此类数据简化为一组少量的具有理论针对性的序次过程的指标集,然后再用任何恰当的分析技术分析这些指标集。

历史演进

序贯分析的发展受到1960年代人类和动物行为学【Ethology】的影响和激发,而且很大程度上由对动物行为和人类发展感兴趣的研究者主导。与行为学方法一致,序贯分析通常也被视为一种归纳事业:观测者将创建一份习性谱(ethogram)(一份关于一个动物可能发生的行为的清单),然后记录行为出现的顺序。这类数据由事件序次构成:如果习性谱(或编码方案)由五个码(如A,B,C,D,E)组成,一个数据序次可能是A B C B C D E C E C D A,诸如此类。其观念是,如果有充足的数据和正确的分析,模式就会突显出来。

由此,研究者创建了交叉分类表:行代表给定的或当前的码(滞后0),列代表目标的或下一个码(滞后1),序次中码的接续对被一一统计于表中(第一个c_1c_2,然后c_2c_3,c_3c_4等,其中c_i代表第i个位置的码),同时将观测频次超过期望频次的单元格标记出来。观测频次超过期望频次的程度能用调整的z分数——一个常见的列联表【Contingency Table】统计量——予以量度。当码很多且没有针对性的假设限制所要检验的码时,这种方法会变得极不易处理。例如,在1960年代中期,计算机革命之前的一篇有影响的文章中,斯图尔特·阿尔特曼描述到,为了寻找出模式,他的手和膝盖下满是矩阵的纸张,这些纸张每张上都绘制有数以百计的行和列的矩阵,它们交叠在一起,铺满了整个地板。

滞后的事件序贯

就事件序贯而言,通常滞后1的表格被

证实是最有用的,但是也可以考虑将滞后2、滞后3等(列代表现有行为之后的第二序次的行为、第三序次的行为等)排在当前序次后,即吉恩·萨基特(Sackett,1979)所谓的滞后序贯分析(lag-sequential analysis)。例如,如果在A的滞后1的位置上,B显著多地出现,同时,在A的滞后2的位置上,C显著多地出现,那么这表示ABC是一个常见的序次,如果在B的滞后1的位置上,C显著多地出现,这就更是如此。当事件的持续期没有太多变化且分析的首要考虑是事件的顺序而不是时间变化时,这类分析最有效。

计时的序贯:一个例子

前文的例子假定事件按顺序进行编码,但是记录它们的开始或结束的时间。当有可用的计时信息时——并且当前的数字革命使得计时信息更容易记录了——序次由此获得了新的分析机会,以无法用于事件序次的方式获得呈现和分析。假定以精确到秒的精度记录时间。那么,计时的序次可以由任意数量的互斥的或同时发生的行为构成,同时在创建列联表时,使用时间单位,而不是事件次数作为计算单位。这样做可能很有用。通常,研究者想要知道相对于另一种行为,某种行为是否会在一个特定的时间中出现,并且他们并不特别关心它的滞后位置,亦即他们并不关心其他行为是否会介入。

例如,德克内、亚当森和巴克曼(Deckner, Adamson, & Bakeman, 2003)想知道,母亲与她们初学走路的孩子的节奏发声(rhythmic vocalizations)是否能相互匹配,并因此针对母亲和孩子的节奏发声的开始和结束时间进行了编码。表1显示了一对母子的匹配结果。

表1 学步期儿童节奏发声的开始时间是否发生在母亲开始发声后的5秒内的交叉表

母亲开始发声后的5秒内	学步期儿童的开始时间		总计
	是	否	
是	11	189	200
否	29	97	1 000
总计	40	1 160	1 200

注:这一对母子的观测期为1 200秒,亦即20分钟;计算单位为秒。

对于表中的行而言,时段被划入一个5秒的时窗内,即由母亲开始一次节奏发声或不发声的那一秒开始计算的5秒持续期;而对于表中的列而言,时段是按每一秒划分的,即学步期儿童开始一次节奏发声或不发声的那一秒。就此一2×2的表而言,一个有效的综括是,计算学步期儿童在他或她的母亲开始发声的5秒内开始节奏发声的概率,即0.058 2∶1或1∶17.2,或者相应的不发声的概率,即0.029 9∶1或1∶33.5。于是,优势比【Odds Ratio】(一个在流行病学领域比其他社会科学领域有更多应用的统计量)为1.95。

优势比在两个方面是有用的:首先,可以用它来描述一种行为在某一个特定的时窗内发生的概率同其不发生的概率相比有多大。其次,优势比的自然对数的取值范围介于负无穷到正无穷之间;当取值为0时,表示没有影响,是一个极适于标准统计分析的分(优势比本身不具有这种特性,它的取值范围介于0到正无穷;当取值为1时,表示没有影响)。据此,德克内等人可以说,24个月大的女孩比同龄男孩或者18个月大的男孩或女孩更可能跟随她们母亲的节奏发声。

总之,德克内等人的研究提供了如何用

序次分析处理计时序次的优秀示例。首先，记录事件开始和结束的时间；其次，用一种计算机程序（Bakeman & Quera，1995）计数时段，同时针对单独个案，计算相应的序次过程的指标（在这里即一个优势比）；最后，一个标准的统计技术（这里是混合方差分析【Analysis of Variance】模型）可用于分析序次分（这里是优势比的自然对数）。

——Roger Bakeman

（马妍译　赵锋校）

参考文献

Bakeman，R.，& Gottman，J. M.（1997）. *Observing interaction：An introduction to sequential analysis*（2nd ed.）. New York：Cambridge University Press.

Bakeman，R.，& Quera，V.（1995）. *Analyzing interaction：Sequential analysis with SDIS and GSEQ*. New York：Cambridge University Press.

Deckner，D. F.，Adamson，L. B.，& Bakeman，R.（2003）. Rhythm in mother-infant interactions. *Infancy*，*4*，201-217.

Sackett，G. P.（1979）. The lag sequential analysis of contingency and cyclicity in behavioral interaction research. In J. D. Osofsky（Ed.），*Handbook of infant development*（1st ed.，pp. 623-649）. New York：Wiley.

顺序抽样（Sequential Sampling）

顺序抽样先要抽取多个样本【Sample】，再一个接一个地实施调查，直至累计的样本满足调查目标为止。时常有的是，很难事先知道一个具体的抽样调查的回答率和每个访问的成本，特别是针对特殊总体【Population】或当回答模式不断变化时。在这些情况下，很难准确规划出要满足研究目标而必须抽取的样本量，如要获得完成回答所需的样本数目、数据收集所需的预算。顺序抽样依赖的原则是：从同一研究总体中抽取的多个概率【Probability】样本，就像一次性抽取的单一样本一样，可以被组合在一起，并做统计估计。顺序抽样允许研究者在不过渡地损坏概率选样的标准或减少无回答【Nonresponse】的情况下，精确地控制一项调查获得的完全回答的数量（Kish，1965）。

顺序抽样通常抽取三个样本：具有一半样本量的一个初始样本，而该样本量在乐观期待下能满足研究目标；从同一总体中抽取的、同第一次取样量相同的另一个样本；以及一个最终的补充样本，也取自同一研究总体。乐观期待下能满足研究目标的样本量可根据某个较为保守的回答率来计算。一般而言，这个样本量等于用调查所需的完成回答的数据除以保守的回答率，并以这个样本量作为数据收集的乐观样本量（即 $n_0 = n/r$，其中 n_0 是数据收集所需的乐观样本量，n 是所需的完成回答数目，r 是对回答率的一个乐观估计）。在实践中，第一个样本（$n_0/2$）通过使用调查设计所指定的所有回收程序而予以完全实施。当第一个样本的数据收集几乎完成时，开始抽出第二个样本，数据收集按计划进行。研究者根据第一个样本计算无回答率和每个访谈的成本。用这些估计量计算第三个样本量，再从同一调查总体中抽取第三个样本；然后，根据调查设计，对它们进行数据收集。顺序抽样允许研究者避开障碍，以免发生不确定的结果。

尽管顺序抽样有明显的益处，但是它增加了田野工作的花费，也让调查实施更加困

难。顺序抽样的一个常见问题是:当第一个样本的结果正在被列表计算时,田野工作会中断。在上述的若干步骤中,第二个样本的使用能极大地缩短田野操作的中断。此外,现在访谈员通常直接将数据输入自动化的档案中,减少了确定最终样本的样本量所需的计算时间。

尽管上面描述的三个样本,即顺序样本,一般适合于获取横截面调查数据,而这一技术已经在以下两个方面得到了扩展。第一,电话调查实施者常从商业抽样公司得到多个样本或重复样本,并将它们作为序列样本来实施调查。通常的做法是,在调查开始时,对若干重复样本实施调查,而当回答率从第一批样本中估出后,为了收集数据,再对额外的样本实施调查。因为商业抽样公司会不断地更新可用的电话号码档案,并对每个独立的样本索取设立费,所以许多电话调查单位会例行地在若干重复样本中得到额外的样本,从而一面控制调查成本,一面最大限度地控制数据收集的不确定性。

第二,顺序抽样技术还被用于追踪民意测验或滚动的样本调查,而这些调查既容许对研究总体做横截面的估计,也容许做时间序列【Time-Series】的估计。追踪的民意调查主要由候选人、媒体和学者用于政治选举(John, 1989; Moore, 1999; Rhee, 1996)。追踪的民意调查可以估计出一个候选人目前的支持度,以及支持度的历时变化(Green, Gerber, & De Boef, 1999)。通常,每天使用一个重复样本来收集数据,而估计候选人的支持度由3天或5天的移动平均数【Moving Average】来估计。滚动的样本调查使用同样的技术来调查更广泛的社会科学研究问题,如意见的动态变化(Henry & Gordon, 2001)。

——Gary T. Henry
(王玥译　赵锋校)

参考文献

Green, D. P., Gerber, A. S., & De Boef, S. L. (1999). Tracking opinion over time: A method for reducing sampling error. *Public Opinion Quarterly*, *63*, 178-192.

Henry, G. T., & Gordon, C. S. (2001). Tracking issue attention: Specifying the dynamics of the public agenda. *Public Opinion Quarterly*, *65*(2), 157-177.

John, K. E. (1989). The polls—A report. *Public Opinion Quarterly*, *53*, 590-605.

Kish, L. (1965). *Survey sampling*. New York: Wiley.

Moore, D. W. (1999, June/July). Daily tracking polls: Too much "noise" or revealed insights? *Public Perspective*, pp. 27-31.

Rhee, J. W. (1996). How polls drive campaign coverage. *Political Communication*, *13*, 213-229.

序列相关(Serial Correlation)

当一个时间序列【Time-Series】过程的接续取值给出非零的协方差【Covariance】时,即存在着序列相关。术语序列相关的用法常等同于自相关,而后者指更一般的同变量自身或一段时期后同自身相关的情形,比如,今天总统的支持率同昨天他的支持率是相关的(序列相关),或者在一定空间内,如帮派成员间在他们的居住地同某个城市低收入区的物理距离上存在相关关系(空间【Spatial】相关)。空间相关的普及性不及序列相关,并且空间相关在教科书关于自相关的部分很少有讨论。序列相关是时间序列

过程的一个共同特征，而且它对于动态模型设定【Specification】具有若干含义。在一个回归模型误差项上，序列（或空间）相关的存在违背了高斯-马尔科夫【Gauss-Markov】回归假定，同时使得源于最小二乘估计的推断无效。

如果一个时间序列过程在时点 t 的取值是它自己先前若干（p）个取值的一个线性【Linear】函数，这一时间序列过程即是序列相关的：

$$y_t = a + b_1 y_{t-1} + b_2 y_{t-2} + \cdots + b_p y_{t-p} + \mu_t$$

如果 y_t 仅同 y_{t-1} 相关，那么我们说，有一个一阶序列相关或一阶自相关，通常记作 ρ_1。如果 y_t 同 y_{t-1} 和 y_{t-2} 都相关，那么我们有一个二阶自相关，记作 ρ_2，诸如此类。自相关系数 ρ 给出了相关【Correlation】的强度。序列相关并不依赖于时间周期自身，仅依赖于时间周期间的间隔数量（即滞后【Lag】的长度），而相关估计的就是两个时点间的关系。y_t 和 y_t 之间的相关是完美的或者等于 1（$\rho_0 = 1$）。估得的序列相关可以告诉我们 y_t 过程中内在固有的时间依赖的性质和程度。就一阶序列相关，也是对最常见的一种序列相关而言，接续的两个滞后间的相关将以某个确定的速率 ρ_1 衰减，而大的相关系数表明更强的时间依赖及更慢的衰减。

如果一个序列是协方差平稳的（stationary）［均值、方差和滞后的协方差（lagged covariances）不随时间变化］，那么可以将 y_t 和 y_{t-p} 之间的自相关定义为 y_t 和 y_{t-p} 间的协方差除以 y_t 的自协方差（方差）。此外，对于一个平稳的序列，相关系数的取值范围介于−1.0～1.0，较大的数值表示较强的跨时依赖性。

例子

通常，社会过程会展现出确向的（positive）序列相关，而非不定向的（negative）序列相关：接续的观测值倾向于朝同一方向移动——或者向上，或者向下。思考以下的例子。

（1）政府或机构的预算以以往的预算为基础，而不是一年一更新的，因此这一过程中的预算变化是递增的。两个临近期的预算相关将是强的和正的。

（2）总计的党派平衡（如民主党的比例）倾向于缓慢地朝某一方向移动，而这或许是因为政治体系的某些特征能保持党派成员的平衡，使其不随时间发生剧烈波动。这将在党派平衡上产生强的确向序列相关。

（3）由于女性平均受教育程度的上升，在专业或管理岗位上工作的女性比例趋于历时的缓慢增加，从而在两个邻近的比例值之间出现强的确向相关。

相反，如果某个过程的当前取值同它的过去取值的关联，在上文 b 取值的正负号上出现震荡时，即出现不定向序列相关。如果一个过程会随时间对过去实现中出现的某些错误发生调整，那么这一过程将造出不定向相关，即在一段时期中，过程朝向上，在下一个时期过程朝向下。

含义

序列相关对于建模【Modeling】有两项含义：其一，分析者在选择一个模型设定时，必须对诸种时间序列的动力过程保持警觉。我们所关注的过程中存在的序列相关可通过将滞后的因变量和可能的自变量直接纳入过程的回归模型而予以建模。其二，分析者也必须意识到在一个动态回归模型中，误差过程自身可能也是序列相关的；亦即即使在模型中已经直接考虑了动力过程，一个时间周期中的误差也会以一种系统的方式同过去的误差相关联。最典型的情况是，预测

【Prediction】误差系统性地太高或太低，从而在模型的误差项上，出现确向序列相关。一个回归模型误差项上的序列相关违背了误差不相关的回归假设，同时，由于估得的标准误【Standard Error】是有偏的，致使统计推论变得无效。偏差的方向随相关系数的符号变化，即正的相关系数会缩小标准误，负的相关系数会放大它们。由于社会科学过程倾向于呈现正相关，通常方差将被低估（我们会放大对估计值的精度的估计），从而可能使分析者去拒绝一个真的虚无假设。此外，在这种情况下，R^2 会被高估。更普遍的情况是，如果存在着序列相关的误差项，那么 t 检验和 F 检验就不再有效而且还会产生误导（Gujarati, 2002）。

鉴于从误差序列相关的模型中获取推断【Inference】有较高的成本，在估计涉及时间序列数据的回归【Regression】模型时，恰当的做法包括针对误差过程中可能存在的序列相关，做估后诊断检验（postestimation diagnostic tests）。这类检验中最常见的是针对一阶自相关的德宾-瓦特逊【Durbin-Watson】检验。

误差项上出现序列相关的原因可能有几种。首先，它的出现可能是由于动态误设【Misspecification】。这里又有几种情况，如：

（1）如果变量的行为方式具有时间依赖性，没能将这些变量纳入模型会导致回归误差项上出现序列相关；亦即模型的误差项中将包括遗漏的时间-因变量。例如，如果我们在布什总统的支持率模型中遗漏了 9 月 11 日，那么将系统地低估了他在那段时期之后的支持率。在这种情况下，误差项上会出现强的正自相关。

（2）不正确的函数【Function】形式可能产生序列相关。如果在我们的模型中 X 和 Y 呈线性关系，而实际上，Y 线性地与 X 的平方相关，那么就会系统地高估一系列连续的 Y 值，然后再连贯地低估另一系列的 Y 值。

（3）如果因变量是惯性的，那么遗漏一个滞后项将产生序列相关。

纠正动态误设的最好办法是更正模型中动力过程的设定。

此外，有些隐藏的误差本身可能即是序列相关的。比如，由数据带来的问题。数据操纵（data manipulation）——插入缺失值【Interpolation】、通过平滑【Smoothing】时间序列处理数据中的随机【Random】波动（对接续的观测值取平均值）或者取几个历时的观测值的和——都可能导致模型的扰动项上出现系统的模式。

随时间推移汇总，这些都可能导致模型中的干扰出现系统模式。在这些情况下，欲得到有效的统计推论，就需要如（可行的）广义最小二乘【Generalized Least Squares, GLS】法之类的补救措施。

其他问题

上文的讨论集中于平稳的时间序列过程背景下的序列相关。平稳过程代表了时间序列过程的一个重要类别，而这类过程是稳定的，因为就所有的时间周期而言，它有一个恒定的均值、方差和协方差。对于稳定的时间序列，序列相关的模式显现为迅速的衰减，即时间序列中近期的值比远期的值与当前的值有更强的相关，同时，相关强度上的衰减率是可预测的和快速的。但是序列相关可以采取多种形式。如果序列是不平稳的，两个接续的滞后间的序列相关系数可能等于或者大于 1。在序列相关系数等于 1 的情况下，过去的效应不会衰减，而是随时间的推移发生累积。相关系数没有出现衰减的序列相关过程被称为单位根（unit root）或者随机游走【Random Walk】过程。那些在更长的两个滞后之间显现出更强的相关关系的序列相关模式被

称为爆裂过程(explosive processes)。这两类过程都是堆积过程(integrated processes)的实例。此外,序列相关虽然可能历时地衰减,但是衰减得很慢以至于效应持续存在于一系列较长的滞后中,这类过程则被称为轻微堆积。由于传统的概率【Probability】理论不能用于任一类型的非平稳过程,因此,在进行假设检验及就社会过程做统计推论之前,必须通过某种方式的变换,将此类序列变成稳定的。

——Suzanna De Boef

(马妍译 赵锋校)

*也可参见自回归求和移动平均模型【Autoregressive Integrated Moving Average, ARIMA】。

参考文献

Gujarati, D. N. (2002). *Basic econometrics* (4th ed.). New York: McGraw-Hill.

Shapiro-Wilk 检验(Shapiro-Wilk Test)

Shapiro-Wilk 检验(通常称为"*W* 检验")是一种用以检验一个随机变量【Random Variable】是否为正态分布【Normal Distribution】的方法。直观地说,该检验统计量即标准的正序统计量(standard normal-order statistics)的线性【Linear】组合的平方同样本方差的比值(或者,换言之,即样本值同它们标准化的、期望的标准正序统计量之间的皮尔逊相关系数【Pearson's Correlation Coefficient】的平方)。计算公式如下:

$$W = \frac{\left(\sum_{i=1}^{N} a_i X_i\right)^2}{\sum_{i=1}^{N} (X_i - \bar{X})^2} \qquad (1)$$

其中,X_i 是样本数据 **X**(从最小值到最大值)的排序值,a_i 是带有方差-协方差矩阵 $\boldsymbol{V}$ 的标准化的正序统计量 $\boldsymbol{m}' = (m_1 \cdots m_N)$,由此 $\boldsymbol{a}' = (a_1 \cdots a_N) = \boldsymbol{m}'\boldsymbol{V}^{-1}[(\boldsymbol{m}'\boldsymbol{V}^{-1})(\boldsymbol{V}^{-1}\boldsymbol{m})]^{-\frac{1}{2}}$ 是非对称的,同时,根据概念有 $\boldsymbol{a}'\boldsymbol{a} = 1$。在正态性的虚无假设前提下,*W* 渐进收敛于夏皮罗和威尔克(Shapiro & Wilk, 1965)给出的分布。低的检验统计量取值支持拒绝数据是正态分布的虚无假设【Null Hypothesis】。

原初的 Shapiro-Wilk 检验是针对样本规模为 3~50 的样本设计的;最近通过修正,该检验扩展了其适用性,可用于检验样本量为 2 000 的样本(参见 Royston, 1993 了解概况),而 Shapiro-Francia *W'* 检验对样本量为 5 000 的样本也是有效的。该检验是尺度不变和原点不变的,同时,对非正态性的背离也非常稳健;基于这些原因,它被视为其他绝大部分正态性检验(如 Bera-Jarque 检验、Kolmogorov-Smirnov 检验【Kolmogorov-Smirnov Test】的 Lilliefors 修正)的卓越替代。

例如,考虑吉尔提供的来自 81 个国家总计的 IQ 数据(Gill, 2002, p.103)。测试工具设计的回答均值为 100,方差为 15;相比而言,样本数据的均值为 88,方差为 12。Shapiro-Wilk 统计量 *W* 等于 0.944,而 Shapiro-Francia *W'* 等于 0.950;两个值都能让我们在 $p<0.01$ 的水平上拒绝正态性的虚无假设。

——Christopher Zorn

(马妍译 赵锋校)

参考文献

Gill, J. (2002). *Bayesian methods: A social and behavioral sciences approach*. London: Chapman & Hall.

Royston, J. P. (1993). A toolkit for testing for non-normality in complete and censored samples. *The Statistician*, *42*, 37-43.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*, 591-611.

显示卡片(Show Card)

一张显示卡片是一个视觉工具,常用于标准化的面对面的定量的调查访谈【Interviewing】。之所以这样称呼它,是因为传统上它由卡纸做成,常用来向受访者展示特定信息。通常,卡片上的信息由对一个问题的一组可能的答案列表构成,同时,被访者被要求从中选择一个或多个合适的答案。使用卡片的主要理由是要克服如下困难:如果访问员只是读出选项,而没有提供视觉帮助,许多被访者就只能在他们的头脑中保留回答选项。

在缺乏一张显示卡片的情况下,这些困难有时会导致临近效应(recency effects),即由于早先的选项记不清了,被访者可能倾向于选择最后一个或后面的若干选项之一(即最近被读出来的选项)。因此,如果有许多个回答选项时(例如超过三个或四个),就可以使用卡片,或者如果选项很长或很复杂(如用来对某一话题加以描述的一组可能的观点)。然而,显示卡片不能克服同回答选项的预先列表相关的所有问题。在某些情况下,回答可能易受首位效应(primacy effects)的影响,即被访者倾向于选择第一个选项,以节省认知负担和读完整个卡片的时间。

一张显示卡片的另一个用途是作为开放性问题【Open-Ended Question】的一个记忆触发器。例如,一个问题问被访者在过去7天里是否从事过任何志愿活动,而这个问题就可以使用卡片列出一些志愿活动的例子,从而给出研究者有兴趣了解的活动种类。这个技术可以帮助促进对那些可能会被忽略的回答的记忆,但是它也可能导致受访者只关注列出来的选项。换言之,受访者可能将列表视作穷尽的选项,而不只是一些例子。

显示卡片也常用于展示那些不适合语言交流的物品。例如,可能给受访者看一个报纸广告,问他是否可以回忆出看到的内容。实际上,用这种方式展示的物品,与那些文本的记忆触发器之间并没有非常大的不同。它可能促进认知,而如果访问员只是用语言描述这些广告,这些认知就可能不会出现,但是另一方面,它也可能触发错误的认知。

能够使用显示卡片(和其他视觉辅助)经常被视作面对调查访谈优于电话调查访谈的一个重要优势。然而,显示卡片也不是没有问题。与口头提问不同,他们需要受访者有一定程度的阅读能力,而这可能成为回答偏倚【Response Bias】的来源,也存在着首位效应的风险。需要注意的是,这些问题也适用于自填式调查【Self Completion Questionnaire】。使用卡片的逻辑方面也是很重要的。制作卡片必须可以轻松地从一张轮换传到下一张,还必须有(以及能保

持)正确的顺序(活页卡片不是一个好点子)。卡片应该有清楚的标注,还应当给访谈员提供清楚的使用指示,即哪一张卡片用在访谈的哪一个问题上(如"现在看卡片C……")。

——Peter Lynn

(王玥译　赵锋校)

符号检验(Sign Test)

符号检验最简单的形式是对两类结果的概率的比较。同样的程序也可用来推论一个差值总体的均值。

设想有一个包含22个登记选民的随机样本【Random Sample】,其中他们被询问在即将到来的选举中倾向于投票给候选人A还是候选人B。7名选民说他们喜欢候选人A,12个偏好候选人B,其余3人则不确定。这是否表明,相对于候选人A,对候选人B有显著的偏爱?我们可以忽略3个未作决定的选民,并假定他们有相等的可能在最终投票中选择任意一位候选人,从而仅考虑剩下的$n=19$位选民,以及其中对候选人B的12次"成功"。检验的基础是样本中的成功数S。如果选民对候选人的偏好是同等的,就每一位投票而言,候选人B获得成功的概率将是0.5,候选人A的情形也一样。对候选人B而言,$S=12$或获得更多成功数的概率可根据二项分布【Binomial Distribution】计算如下:

$$P(S \geqslant 12)=\sum_{S=12}^{19}\binom{19}{S}(0.5)^{S}(0.5)^{19-S}$$
$$=0.179\ 6$$

这个值并不是非常小,因此,我们认为这个样本没有提供足够的证据,让我们推论说,选民显著地偏爱候选人B。对于更大的样本量,可根据二项分布的正态【Normal】近似(有连续校正)计算检验统计量:

$$Z=\frac{S-0.5-\dfrac{n}{2}}{0.5\sqrt{n}}=0.92$$

这一正态近似给出$P=0.178\ 8$,而这个值与确切值极为接近。

符号检验的另一种运用是:设若我们有一个取自某个二元分布,包含n对观测值的一个随机样本(X_1,Y_1),(X_2,Y_2),…,(X_n,Y_n)。我们想要推论总体中X_S和Y_S的差值之均值。第一步取每一对观测值X和Y的差值,即$D_i=X_i-Y_i$。因为X和Y总体均值之差等于差值总体$D=X-Y$的均值μ_D,所以有虚无假设$H_0:\mu_D=0$。

为了进行符号检验,我们先求得差值$D_1,D_2,\cdots,D_n$,然后记下这些差值中取正号的数目S。如果备择假设是$H_1:\mu_D>0$,那么S值很大的话,检验就会拒绝虚无假设,或者等价地说,P值将落在成功概率$p=0.5$的二项分布的右尾部。符号检验的此类应用是二项检验【Binomial Test】的一个特例,而二项检验的虚无假设是$p=p_0$,备择假设是$p>p_0$。

例如,设想随机选择出10对夫妇,并判断丈夫和妻子的评级之间的关系,评级的内容涉及冷冻的新月形比萨的口味、便利性和价格。答案是一个0(极糟的产品)到100(极好的产品)的量度上的得分。数据见表1。丈夫对该产品的喜爱甚于妻子吗?

表 1 偏好数据

夫妇	1	2	3	4	5	6	7	8	9	10
丈夫	76	83	60	90	100	55	62	69	75	43
妻子	71	81	62	84	65	44	61	67	75	40
差值	+	+	−	+	+	+	+	+	0	+

我们忽略其中有相同评分的夫妇。在剩余的 9 对夫妇中，有 8 对丈夫比他们的妻子更喜欢该产品。在等偏好虚无假设的前提下，这种情况发生的概率是 $P(S \geqslant 8) = 0.0195$。由于 p 值如此小，我们拒绝虚无假设，并推论丈夫的偏好评分均值大于妻子的评分均值。

——Jean D. Gibbons

（马妍译　赵锋校）

显著性水平(Significance Level)

参见 Alpha（α），检验的显著水平【Alpha, Significance Level of a Test】。

显著性检验(Significance Testing)

显著性检验提供了客观的规则，以确定一位研究者的假设【Hypothesis】是否得到了数据的支持。对于客观性的需求源于这样一个事实：假设与目标总体【Population】的值有关，而数据来源于在代表总体的信度上有所不同的诸样本【Sample】。例如，兴趣总体可能是大学年龄的人群，而样本可能是在数据收集时可用的一个特定班级。当然，这可能不是一个理想的随机抽样【Random Sampling】，但却是一个经常在社会科学研究中使用的方便样本。

（在决定从一个方便样本获得的数据的结论在多大程度上可推论至样本之外时，就需要运用特别的统计考量。例如，根据一个大学生的方便样本获得的视觉感知的研究结果应当是可以做一般性推论的）为了推论一个特定变量【Variable】对行为的影响，研究者必须在获取和记录数据时遵循好的研究设计原则。否则，对数据进行再多的统计"按摩"也无法挽救一个设计不佳的研究。

即使数据源于一个设计良好的研究，但只说明两组取值之间有差异是不够的。研究者必须证明两组行为之间的差值是可信的【Reliability】，此即，它们大于那些可归之于机会的差值。

机会或（更正式的）机会抽样效应是指预先存在于样本个体之间的差异，而这些差异同研究中如何将其归类或处理都没有关系。在一项关于人类行为的研究中，机会抽样效应可能是实质性的，即受试者先天的基因遗传和环境影响过程是不在研究者控制范围之内的。为了消除分配受试者带来的所有系统偏倚，即根据未受控制的因素，将受试者分入不同的条件，从而产生系统偏倚，研究者通常使用随机化方法，将受试者分入不同的条件，即所有受试者都具有同等机会被分入任意一组条件。相比于从某个未精确定义的总体中选取受试者而言，这是更为严格的随机性【Randominess】的应用。

然而，就算使用了随机性，一个像样本均值【Mean】这样的量度也会随样本的不同

而有相当大的变化,即便从中取样的总体是等价的。变差可归之于这一事实,即不同的个体进入了不同的样本,同时每个个体都具有唯一的特征。以致不可避免地,样本之间会有变化。当样本量较小,且少量有差异的取值就会扭曲结果时,尤其如此。后面,我们会设计一个具体的例子,其中会发现样本量在显著性检验中发挥关键作用。事先说明这个示例的结论,样本量越小,为了证明组间均值差是一个统计显著的差值,就要求这个差值本身更大。

研究者确定一个差值是否是"统计显著的"正式过程,亦即,这是否表示该差值不只是机会变差,被称作假设检验【Hypothesis Testing】。

假设检验

社会科学中的绝大部分研究问题可归结为两个竞争性假设。这两个竞争性假设被称为虚无假设和备择假设。虚无假设【Null Hypothesis,H_0】是最初假定为真的假设。大多数情况下,H_0 假定没有差异(如实验处理没有效应);而备择假设【Alternative Hypothesis,H_a】一般断言实验条件之间存在差异。虚无假设通常与实验的实际预测不一致。它更像一个研究者希望质疑它的"稻草人"。然而,因为虚无假设陈述的是相等(如实验组和控制组的总体均值的差为 0),而备择假设陈述的是不相等(如均值差不等于 0),但是,只有 H_0 可以直接检验,这也就是我们从虚无假设开始的原因,即便它不是真正的研究假设。

关于虚无假设的示例包括如下陈述:一种药没有效果;无论是单独的还是在小组内的,个体都将作出同样的决定;收入和政党归属之间没有关系;男性和女性同工同酬。为了断言一个效应或一种关系【Relationship】的存在,研究者必须首先拒绝 H_0 或使其无效(这就是为什么称它为"空"假设的原因),即通过给出同没有效应的假设不一致的数据,使虚无假设失效。因此,人们使用了间接证明(indirect proof)或间接支持(indirect support)这类术语。

在检验这些假设的过程中,特定统计检验(如 t 检验【t-Test】、方差分析【Analysis of Variance】、回归【Regression】、卡方检验【Chi-Square Test】)的使用取决于用来生成被检验数据的特征。这包括所设计的研究是用于检验实验条件之间的差值,还是衡量不同组的观测值之间的关联【Association】强度、所收集数据的类型(如数值型数据或频数),以及实验设计【Experimental Design】中因子(变量)的数量。引自莱文(Levin,1999)的流程图(图 1),例示了基于上述考量的检验选择路径(也请参见 Howell,2002;Tabachnick & Fidell,2001)。

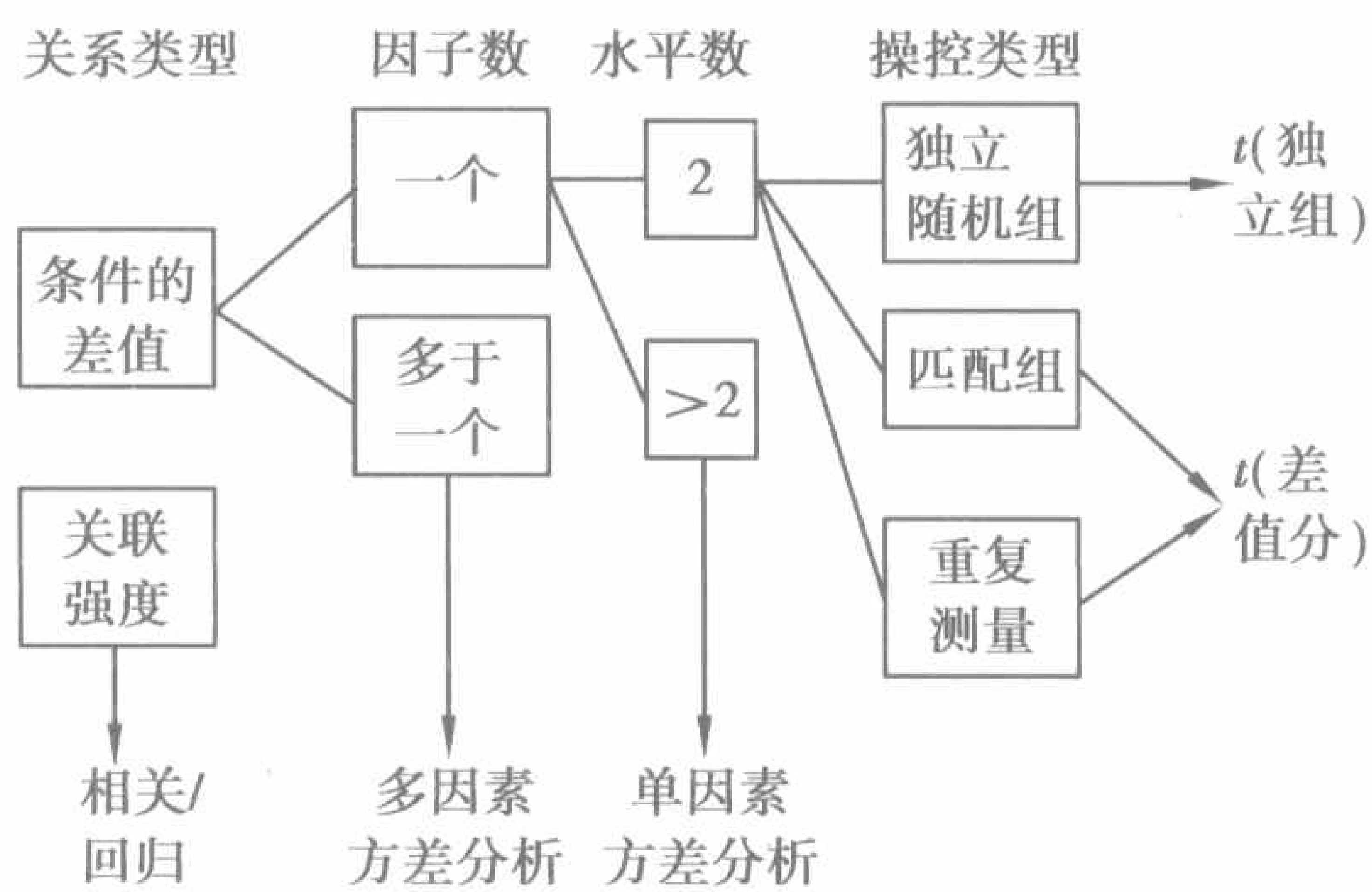


图 1 检验的选择

资料来源:Levin,1999。

无论特定的统计检验是否适合于一个给定的数据集,逻辑总是相似的。根据样本数据中实际观测到的诸条件之间的差值或关联度,计算一个量度。再将这个量度同虚无假设为真时的期望值相比较,而所谓虚无假设为真,即诸条件间不存在差异,或观测集之间没有关联,仅有机会因素影响数据。

观测到的差异超过基于机会预期到的差异越多,就越不可能将差值归之于机会,也就越不可能将它们归之于其他因素。在

一个设计良好的实验研究中，那些“其他因素”可能是自变量【Independent Variable】的操控对因变量【Dependent Variable】的效应。而在一项设计良好的观测【Observational】研究中，“其他因素”可能是不同类型人群之间的差异，如民主党人和共和党人在给政治候选人捐款数额上的行为差异。

现代计算方法容许我们计算一个给定的数据可能仅是机会因素的后果的概率【Probability】（也称为 p 值【p Value】），亦即虚无假设实际为真的情况下，一个给定数据出现的概率。这个概率越小，虚无假设就越不可能为真。因此，当 p 值较小时，拒绝 H_0 而支持 H_a；而当 p 值较大时，则保留 H_0。但是哪些特质说明 p 值的“小”和“大”？我们之前说过，显著性检验为检验假设提供了客观规则。

在此，仍要用到一些主观性。无疑，如果 $p<0.01$，那么我们希望拒绝 H_0，因为，如果虚无假设为真，那么在 100 次中，得到这样一个数据的机会不到 1 次。同样的，如果 $p>0.25$，那么会保留 H_0，因为，结果很可能是由于机会，那么在这两个值之间的 p 值又如何呢？

这时需要考虑检验统计假设过程中可能犯的推论错误的类型。当 H_0 为真时，你却拒绝了它而支持 H_a；或者当 H_0 为假时，你却保留了它，都会犯错误。按照惯例，这两类错误分别被称为第一类错误【Type Ⅰ Error】和第二类错误【Type Ⅱ Error】。第一类错误的概率（称为 Alpha，记作 α）和第二类错误的概率（称为 Beta，记作 β）成反向相关。如果为拒绝 H_0 设定了一个严格的标准，那么 H_0 为真时它几乎不会被拒绝；但是，由于一个低水平的 α 和一个相对高水平的 β，当 H_0 为假时，就倾向于保留它。为拒绝 H_0 设定一个宽松的标准，就会出现相反的情况。

这个标准被称为显著性水平【Level of Significance】，同时，事先既可以将其设定得相当低，以减少犯第一类错误的概率，也可以将它设定得足够高，以减少犯第二类错误的概率。这一决策必须立足于权衡第一类错误和第二类错误的相对严重性。例如，在医学研究中，第一类错误意味着一种无效的药或者无效的治疗被认为有效果；而第二类错误意味着一种有效的药物或有效的治疗被认为没有效果。因此，一位检验一种新药有效性的医学研究者可能希望设定 $\alpha=0.01$，以避免一种未经证实的药物过早地进入市场。在实践中，通常使用一个水平适中的显著性水平，即 $\alpha=0.05$。

一个例子

t 检验最常用来比较两个组或两种条件的均值。虚无假设使两种条件下的总体均值得分相等。换言之，如果能够比较两种条件下整个总体的得分，将不会发现两者的均值有差异，即 $H_0: \mu_1-\mu_2=0$，其中 μ_1 和 μ_2 分别表示两种条件下的总体均值。但是数据来自样本而非整个总体。例如，在一项有关儿童攻击行为的研究中，一个组或一个条件可能是一组频繁观看电视中的暴力画面的样本，而另一组则是一组少量观看的样本。备择假设是两组之间有差异，即 $H_a: \mu_1-\mu_2\neq 0$（如果事先预测了差异的方向，可以设定 H_a 的一个单尾【One-Tailed】样式，如 $\mu_1-\mu_2>0$）。

检验假定，在每种条件中，得分的分布是正态【Normal】的，并且两个分布具有相等的方差。然而，有大量的证据表明，当样本量相等时，这些假定的适度偏离对 t 检验相对地没有影响。在研究中，要针对两个样本的得分计算如下统计量：均值 $\overline{X}_1$ 和 $\overline{X}_2$；标准差【Standard Deviation】s_1 和 s_2；以及样本量 n_1 和 n_2（注意：希腊字母常用来表示总体的值，而拉丁字母则常用来表示样本的值。这是为了明确区分样本和总体。假设是就目标总体陈述的；但它们是用样本数据来检

验的）。当样本量相等时，即 $n_1 = n_2 = n$，可用下列公式计算 t 值：

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{(s_1^2 + s_2^2)/n}}$$

注意，公式中的分子是样本均值实际差异的一个量度，而分母则包含了那些代表机会的影响的变化性的量度。比值越大，结果越有可能代表的不仅仅是机会变差。

接着，根据选定的 α 水平和恰当的自由度【Degrees of Freedom】，在当前情形中，$df = (n_1 - 1) + (n_2 - 1) = 2(n-1)$，将 t 的结果值与 t 统计量表中的临界值【Critical Value】相比较。临界值确定了 t 统计量的分布上单尾或双尾所占面积的比率，即代表了一个等于 α 的区域。这个区域被称作临界区。如果由数据计算得到的 t 的绝对值大于表中的数值，那么决定是拒绝 H_0，同时支持 H_a。

如果计算得到的 t 的绝对值小于表中的值，那么必须保留 H_0 作为一个可行的选项。保留 H_0 的理由是，观测到的差值的量并未足够大，以至于可以推论两种条件是不同的。而拒绝 H_0 的理由是，两个样本均值之间观测到的差值很大，以至于无法将其归因于机会，因此研究者可以推论，两个组的得分确实是不同的。这常被称为一个统计显著【Statistical Signifcance】的差值。

这里提供一个数字实例，假设家长和老师根据一个 1~7 的量表给每个孩子的攻击行为评分，频繁观看组的平均评分是 0.5，轻度观看组的平均评分是 3.6。进一步假定每一组得分的标准差是 2。为了强调样本量的重要性，让我们先设想，这些数据来自一个每组中有 10 个孩子的样本，然后再考虑样本量为 25 的情形。

在样本量为 10 的例子中，t 值是 1.57；由于自由度为 18，所以在传统的 0.05 的显著性水平上，它不显著（没有超过临界值）。在样本量为 25 的例子中，t 值是 2.47，自由度是 48，该 t 值是显著的。因此，只有在后一个例子中，研究者可以得出这样的结论：结果在统计上是显著的，可以拒绝虚无假设，即认为观看电视的类型对孩子攻击性行为无影响；并接受备择假设，即在电视上观看了许多暴力画面的儿童组比在电视上观看较少暴力画面的儿童组会有更多的攻击性行为。

小结

显著性检验【Significance Testing】旨在根据某种概率或可能性的估计，给出一种决策规则，而这种概率或可能性是一组给定的研究结果，完全可归之于在不同的条件下随机变化的机会因素。当估得的仅由于机会得到一组结果的可能性小于一个预先确定的显著水平时，最合理的推论是，结果不能仅归之于机会，同时，决定拒绝没有效应的假设。如果研究设计和实施得当，那么研究可以得出这样的结论：感兴趣的因素（如被操控的自变量）可能影响了行为。

——Irwin P. Levin

（马妍译　赵锋校）

参考文献

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Duxbury.

Levin, I. P. (1999). *Relating statistics and experimental design: An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-125). Thousand Oaks, CA: Sage.

Tabachnick, B. G., & Fidell, L. S. (2001). *Computer-assisted research design and analysis.* Boston: Allyn and Bacon.

简单相加指数(Simple Additive Index)

参见多项测度【Multi-Item Measures】。

简单相关(回归)(Simple Correlation [Regression])

简单相关【Correlation】是一个用来确定变量 X 和变量 Y 之间关系的强度和方向的量度。一个简单相关系数的取值范围为 $-1\sim1$。然而,某些简单相关的最大值(或最小值)不可能达到绝对(即 1 或 -1)。给定有两组观测值:$x_1,x_2,\cdots,x_n$ 和 $y_1,y_2,\cdots,y_n$,都偏离各自的均值,一个一般的简单相关(r_{xy})可以定义如下:

$$\frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2}}$$

或

$$\frac{\mathrm{Cov}(X,Y)}{s_X s_Y}$$

其中,s_X^2 和 s_Y^2 是变量 X 和变量 Y 的方差【Variance】,而 $\mathrm{Cov}(X,Y)$ 是变量 X 和变量 Y 的协方差【Covariance】。一般简单相关的实例包括肯德尔的 Tau 系数【Tau】、斯皮尔曼相关系数【Spearman Correlation Coefficient】和皮尔逊相关系数【Pearson Correlation Coefficient】。

变量 X 和变量 Y 之间的简单相关并不意味着两者之中的任何一个变量是自变量【Independent Variable】或者因变量【Dependent Variable】。工作满意度和工作绩效之间观测到的关系并不表明,满意度高的员工是否有动机取得更好的工作绩效,或者相反,那些表现好的员工是否会因为对自己感到自豪而更满意自己的工作。更进一步,一个简单相关可能或可能不反映两个变量之间的真实关联,因为它们之间的关系可能受一个或几个第三种变量的影响。例如,就一群小学生而言,其学校成绩和其智力间观测到的关系可能受学生的社会经济地位的影响。

假设变量 X 和变量 Y 之间存在线性关系,通过两个回归【Regression】方程 $Y=a_1+b_{y.x}X+e_1$ 和 $X=a_2+b_{x.y}Y+e_2$,我们可以用变量 X 来预测变量 Y,也可以用变量 Y 来预测变量 X。$b_{y.x}$ 和 $b_{x.y}$ 是回归系数【Regression Coefficient】,a_1 和 a_2 是截距【Intercept】,而 e_1 和 e_2 分别是变量 Y 上不能由变量 X 解释掉的残差【Residual】和变量 X 上不能由变量 Y 解释掉的残差。因为 $b_{y.x}=r_{xy}\frac{s_y}{s_x}$,$b_{x.y}=r_{xy}\frac{s_x}{s_y}$,所以 $r_{xy}=\sqrt{b_{y.x}b_{x.y}}$。从概念上讲,在这种情形下,一个简单相关即两个回归系数 $b_{y.x}$ 和 $b_{x.y}$ 的几何均值。此外,r_{xy}^2 代表了使用变量 X 预测变量 Y 或使用变量 Y 预测变量 X,从而解释掉的方差比率。

不同特征的关系可以用不同的简单相关度量。例如,可以用斯皮尔曼相关系数度量一个单调的【Monotonic】关系,用皮尔逊相关系数度量一个线性关系【Linear Relationship】,用 Somers 的 d 系数度量一个非对称关系【Asymmetric Relationship】。一个正向的或负向的单调关系表明,两个

变量朝着同一个方向或者相反的方向变动。一个线性关系是单调性关系的一种特例,表明根据某种线性函数规则(如$Y=a+bX$)X能够相当好地预测Y(或者反之亦然);当变量Y相当好地预测了变量X,而反向的预测不可能时,即出现一个非对称关系。

简单相关可以推广到两个以上的变量,比如复相关系数【Multiple Correlation】或典型相关【Canonical Correlation】。其中,前者估计一个变量(即准则)和一组其他变量(即预测变量)之间的关系;而后者衡量两组变量之间的关系。此外,简单相关可用于任意测量层次【Level of Measurement】的变量。例如,可以估计年龄(定比【Ratio】尺度)和性别(定类【Nominal】尺度)之间的关系,华氏温度(定距【Interval】尺度)和电量消耗(定比尺度)之间的关系,或者字母等级(定序【Ordinal】尺度)和宗教信仰(定类尺度)之间的关系。

——Peter Y. Chen
Autumn D. Krauss
(马妍译　赵锋校)

参考文献

Bobko, P. (1995). *Correlation and regression: Principles and applications for industrial organizational psychology and management.* New York: McGraw-Hill.

Chen, P. Y., & Popovich, P. M. (2002). *Correlation: Parametric and nonparametric measures.* Thousand Oaks, CA: Sage.

Hays, W. L. (1994). *Statistics* (5th ed.). New York: Harcourt Brace.

简单观察(Simple Observation)

简单观察技术可以生成直接的和第一手的研究资料,此即,那些没有通过解释、人工实验的【Experiment】条件、记忆或其他因素产生或染色的资料。观察的资料可能带有观察者所持有的某种观点,或者之后它们可能被用来帮助发展理论。简单观察技术常常在那些还没有发展得很好的研究领域使用,而对这些领域仔细的观察可能产生未来的研究假设【Hypothesis】。

一些陷阱会影响观察研究中的资料收集和资料诠释,例如,罗森塔尔(Rosenthal, 1976)的著作和社会科学方法的教科书(Lee, 2000)中所描述的。第一个陷阱是一个硬性的问题,亦即是否对观察到的行为实施了记录,以及观察是否可以在没有偏差【Bias】、没有研究参与者的同意或研究参与者没有意识到,或没有禁止性的伦理的【Ethical】困难的情况下进行。

第二个陷阱是观察资料的问题,即如果资料不是可为研究使用的标准化的或永久记录的(这样才可以经得起重复的检查和确证【Verification】),那么就可能产生不可靠的资料,例如,当多个见证者对一个事件提供资料,或者当未经训练的人令人可疑地描述那里"发生了什么"时。因此,社会科学中存在许多例子,其中就同一行为或事件给予不同的说明;这些说明非常类似于由竞争的体育团队的支持者给予一场体育赛事的不同说明和描述。

第三个陷阱是观察者差异这一事实。也就是说,除了解释上的偏差问题,观察者间的差异不可避免地导致资料的方差【Variance】。例如,在时间间隔的估计、频次计数或者一个行为或事件可能的客观特征上都会出现观察者间的变差。有时,这个问题可以通过几个判别人或多个个体观察

者给予即时的(和之后平均的)观察来补救,而且其中每个观察者都根据标准化的指令独立地观察,同时,他们既不知道彼此的身份,也不熟悉所做研究的背景或研究目的。

简单观察的第四个陷阱与观察者自身成为被观察事件的隐秘参与者的可能性有关,即有时那些行为被观察的人不知道研究者参与其中。例如罗森汉恩(Rosenhan,1973)安排了一组参与观察者【Participant Observation】("假病人")分别独立地进入12个不同的美国精神病医院,也就是他们是在真实的病人和医院的职员不知情的情况下进入的。除了由于解释偏差或理论偏好带来的更具色彩的信息外,这种方法具有能够就特定环境中的行为,收集到客观的和事实的观察资料。然而,无疑的是,这些操作也会带来伦理的困境,在某种程度上,为了在不受参与者知识和意识的影响下开展生态上有效的观察研究,这是一种不得不承受的代价。

在一个新的或前人未踏进的领域中,观察方法的最大用途是产生纯粹的、第一手的资料,尤其当观察资料有永久记录时,更是如此。这类资料(以及其中产生的可用于未来研究的想法)的信度【Reliability】和解释不仅可以被评估、被验证,还可以被扩展到其他研究中。

——Stewart Page

(王玥译 赵锋校)

参考文献

Lee, R. (2000). *Unobtrusive methods in social research*. Buckingham, UK: Open University Press.

Rosenhan, D. (1973). On being sane in insane places. *Science*, *179*, 250-258.

Rosenthal, R. (1976). *Experimenter effects in behavioral research* (rev. ed.). New York: Irvington.

简单随机抽样(Simple Random Sampling)

简单随机抽样(SRS)是随机抽样【Random Sampling】的一个特例。它通过一种随机机制来选取被调查单位,而每个单位都有相同且独立的被选中的机会。区别SRS和其他随机抽样形式的两个特征是:选中概率的相等和独立。结果要求,在由 N 个单位构成的一个总体中,由 n 个特定单位构成的每一种组合都有相同的机会成为被选中的样本。

从一个充分混合均匀的袋子中抽选小球常被视作SRS的一个很好的类似。抽取彩票号码的机制就设计为等价于SRS。简单随机抽样很少用于社会调查,因为存在着一些常见的能够结合到设计特征中的好理由,如分层(参见分层样本【Stratified Sample】)、分群(参见多阶抽样【Multistage Sampling】),以及不成比例抽样分数(参见抽样分数【Sampling Fraction】)。然而,它可以提供有效的用以比较备选的样本设计的基准,特别是在偏倚问题和估计量【Estimator】的方差【Variance】问题上(参见设计效应【Design Effects】)。

简单随机抽样在抽样教科书和课程中通常有着重的叙述,因为它提供了一种简单条件,从中可以发展出理论和代数。相应的结果可以进一步地扩展到其他的、更具现实性的样本设计中。

——Peter Lynn

(王玥译 赵锋校)

参考文献

Barnett, V. (1974). *Elements of sampling theory*. Sevenoaks, UK: Hodder & Stoughton.

Scheaffer, R. L., Mendenhall, W., & Ott, L. (1990). *Elementary survey sampling*. Boston: PWS-Kent.

模拟(Simulation)

大部分社会科学研究都需要发展或使用某种理论或模型;例如,一种关于认知的理论或一个关于阶层系统的模型。一般来说,这些理论是用文本的形式来陈述的,当然有时理论也可以用方程式的形式表达(如回归分析【Regression Analysis】)。随着1980年代个人计算机变得更加普遍,研究者开始探索用计算机程序表达理论的可能性。这种方式的优势在于社会过程可以用计算机来模拟,同时,在某些情况下,甚至可以在人造的社会系统中开展实验,而这种实验在人类群体中或者不可能开展,或者是不符合伦理标准。

最早是使用纸笔(或用卡在棋盘格上)来尝试模拟社会过程,但是随着强大的计算机、更加有力的计算机语言,以及更多可用数据的出现,把模拟作为一种方法来发展和验证社会理论的兴趣也在增加。

模拟方法论潜在的逻辑是,一个模型即是通过对那些实际存在的社会过程的理论抽象过程建构出来的(如用计算机程序或回归方程的形式)。而后这一模型被用于生成那些同经验数据相比的预期值。统计模型和模拟的最主要区别在于,模拟模型自身可以"运行",以产生结果,而统计模式要使用统计分析程序去生成预期值。正如这个描述所暗示的,模拟倾向于暗含一种实在论【Realist】的科学哲学。

模拟的优势

要建立有用的模拟模型,需要十分清楚地思考它的理论预设。要模拟的每一个关系都需要有精确的设定,同时,每一个参数都需要赋值,否则是不可能完成模拟的。这一原则意味着假设不可能是含糊不清的。这就意味着,这一模型的所有细节对于其他研究者都是潜在的、开放的、可以检验的。然而,这些清晰和精确的要求也有不利的一面。因为复杂社会过程的模拟包含许多参数的估计值,而要获取充分的资料来作出估计是困难的。

模拟的另一个好处是在一些情况下,可以用它来洞察从微观层面的行动中"突显"的宏观层面的现象。例如,对互动个体的模拟,在一个社会尺度上予以考察时,可能揭示出清晰的影响模式。例如,诺瓦克和拉塔内(Nowak & Latané,1994)的模拟展示了一个简单的规则(个人影响他人的态度的方式)是如何产生一个社会水平上的态度变迁,同时,阿克塞尔罗德(Axelrod,1995)的一个模拟显示,政治统治的诸模式如何可能从民族国家所遵循的少量规则中涌现出来。谢林(Schelling,1971)使用一个模拟来说明,当个人所要居住的邻里社区住着的大部分人都是不同民族时,高度的居住隔离将可能出现。

在这个模拟中,人们的住宅以方块的形式呈现在一个大的网格上。每一个网格或

者有一个模拟的住户(或者是“黑”住户,或者是“白”住户)占领,或者还没有住户占领这个网格。当模拟进行时,每一个模拟的住户依次察看八个邻居的住宅方块,看看有多少邻居是与自己相同的种族,多少是其他种族。如果相同种族的数量不够高(例如,少于三个邻居是自己的种族),这个住户可以随意选择搬去那些没有被占领的其他网格方块。然后,下一个住户再考虑自己的邻居,这样一直进行下去,直到每一个住户都对自己的邻居的种族平衡感到满意。

谢林注意到,当模拟达到一个停顿点时,即住户不再想搬家的时候,总是存在着一个同一颜色的邻里家庭的聚合模式。他提出,这个模拟模仿了白人从那些被黑人占领的社区逃离的行为,同时,他从他的模拟实验中观察到,即使当白人同意住在一个黑人邻居占大多数的社区时,聚合现象仍会出现:甚至当住户准备要与其他种族做邻居时,居住隔离仍可能出现。

行动者为基础的模拟

社会模拟的领域已经被一种称为行动者为基础的模拟(agent-based simulation)(也称为多行动者模拟,multiagent simulation)所支配。尽管其他类型的模拟仍在发展之中,如那些以系统动力学模型(使用几组差分方程)为基础的模拟和微模拟【Microsimulation】(通过模拟一个调查样本的衰老,来了解它在未来的特征),但是现在大部分的模拟研究使用行动者。

行动者是设计用来在一个模拟的环境中相对自主地行动的计算机程序(或程序的组成部分)。在模型中,行动者可以代表个体或组织。行动者一般通过程序而被赋予能够去“感知”情境并对情境作出“反应”,能够去追求他们被设定的目标,以及能够与其他行动者互动,例如,给其他行动者发送信息。行动者通常是由面对对象的编程语言创造出来的,并通过使用条件-活动规则的集合而予以构建。

行动者为基础的模型已经被用于考察领导的基础、规范的功能、组织环境改变的意义、土地使用计划限制对人口的影响、语言的进化,以及很多其他议题。可以从期刊《人工社会和社会模拟》(*Journal of Artificial Societies and Social Simulation*)中找到当前研究。

尽管大部分的行动者为基础的模拟被用来模型真实的社会现象,然而也可能去模拟那些并不存在于我们的世界的情境,以理解是否存在关于社会生活的可能性的普遍限制(例如,如果社会成员完全是自利的和理性的,这个社会还可以运行吗?)模拟的连续谱的一端是完全想象出来的社会,另一端是以极大的细节再现具体的环境。

行动者为基础的建模的一个变化样式是用人代替一些或所有的计算机行动者。这就把模型转变成为一种多玩家的计算机游戏,它的价值在于可以让多个玩家学习更多关于社会环境的动力学(例如,商学院学生使用这种类型的游戏来学习商业策略的影响)。这种游戏被称为参与式模拟。

模拟的潜力

尽管计算机模拟被简单地看成另一种呈现社会过程的模型方法,但是它鼓励一种理论视角,即强调突显,强调寻找那些能够产生复杂现象的简单规律,以及强调一种社会发展的进化观点。这一视角与复杂理论联系在一起,而复杂理论是一种试图确定可用于所有系统的一般原理,并显现自主行为(即包括人类社会,也包括生物和物理现象)的理论。

——Nigel Gilbert

(王玥译　赵锋校)

参考文献

Axelrod, R. (1995). A model of the emergence of new political actors. In N. Gilbert & R. Conte (Eds.), *Artificial societies*. London: UCL.

Carley, K., & Prietula, M. (Eds.). (1994). *Computational organization theory*. Hillsdale, NJ: Lawrence Erlbaum.

Epstein, J. M., & Axtell, R. (1996). *Growing artificial societies: Social science from the bottom up*. Cambridge: MIT Press.

Gilbert, N., & Troitzsch, K. G. (1999). *Simulation for the social scientist*. Milton Keynes, UK: Open University Press.

Nowak, A., & Latané, B. (1994). Simulating the emergence of social order from individual behaviour. In N. Gilbert & J. Doran (Eds.), *Simulating societies: The computer simulation of social phenomena* (pp. 63-84). London: UCL.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, *1*, 143-186.

联立方程(Simultaneous Equations)

许多理论设定变量间具有相当复杂的关系,难以用一个单独的方程予以建模。因此,联立方程系统设定了多个方程,其中每个方程都与其他方程关联。各方程间的可能关系包括:因变量【Dependent Variables】间的因果关系,因变量间的双向关系【Reciprocal Relationship】,或者方程误差项间的相关【Correlation】。联立方程模型可被视为结构方程模型【Structural Equation Modeling】的一个子集,即模型仅考虑测得(observed)变量。

联立方程两个主要的类别包括:递归【Recursive】系统,其中影响的方向朝着单一方向前行;非递归【Nonrecursive】系统,其中包括双向因果关系(reciprocal causation)、反馈环路(feedback loops)或误差项相关等情形。联立方程模型可用以下三种方式中的任意一种来设定:一个方程组、一个路径图【Path Diagram】或系数矩阵。

联立方程系统的方程可以设定成如下递归联立方程系统:

$$y_1 = \gamma_{11}x_1 + u_1$$
$$y_2 = \beta_{21}y_1 + \gamma_{21}x_1 + \gamma_{22}x_2 + u_2$$
$$y_3 = \beta_{31}y_1 + \beta_{32}y_2 + \gamma_{32}x_2 + u_3$$

其中,$Cov(u_1, u_2) = Cov(u_1, u_3) = Cov(u_2, u_3) = 0$。

在方程系统中,内生变量【Endogenous Variable】记作 y,外生变量【Exogenous Variable】记作 x。β 表示一个内生变量对另一个内生变量的效应,而 γ 表示一个外生变量对一个内生变量的效应。u 表示方程中的误差项。

完成模型设定后,我们必须确定模型不存在识别问题【Identification Problem】。此即,必须确定模型的参数是否是可估计的。有很多规则和条件可用来判定模型的可识别性,包括零 beta 规则(the null-beta rule),即不含有 beta 矩阵的模型(看上去像不相关的回归模型)是可识别的;递归规则(the recursive rule),即所有递归模型都是可识别的;顺序条件(the order condition)以及秩条件(the rank condition)(Bollen, 1989, pp. 98-103)。里格登(Rigdon, 1995)提供了一种直接的方法来确定多种类型模型的可识别性,而无须依据上述规则。

递归方程系统的估计是直接的,可以用普通最小二乘【Ordinary Least Squares, OLS】法对每一个方程作回归来求解。非递归方

程系统的估计需要在估计某个方程时考虑系统中其他方程的信息。因此,如果将 OLS 回归用于非递归联立方程系统,那么估计的结果可能是有偏的【Biased】和不一致的。此外,有多种替代的估计技术可用于求解非递归联立方程模型,包括二阶最小二乘(two-stage least squares)、三阶最小二乘(three-stage least squares),以及最大似然估计【Maximum Likelihood Estimation】。以上所有这些估计量都可被视为工具变量【Instrumental Variable】估计量。

——Pamela Paxton

(马妍译　赵锋校)

参考文献

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Greene, W. H. (1997). *Econometric analysis* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Heise, D. R. (1975). *Causal analysis*. New York: Wiley.

Rigdon, E. E. (1995). A necessary and sufficient identification rule for structural models estimated in practice. *Multivariate Behavioral Research*, *30*, 359-383.

偏斜的(Skewed)

某一具有偏斜分布的变量其值的多数聚集在变量值的分布【Distribution】的一端。该变量的各种集中趋势【Central Tendency】变量不相等。偏斜的后果可能很严重,特别对普通最小二乘【Ordinary Least Squares, OLS】回归分析而言。

为便于说明,设想一个对民主的测量,有一个从 0(不民主)到 100(完全民主)的量表,而该量表构建的依据是判断国家在符合民主概念的各类别上得分的高低(比如,公民权和自由,行为自由和选举公正)。如果北半球国家被选中,那么结果将显示经济合作与发展组织(OECD)集中在民主得分的高端,从90~100。这将呈现民主分布上的负偏斜,且意味着分布的极值位于量表左侧 90~100 的连续统上。如果南半球国家被选中,由于有更多的威权政权,就会有许多国家得分在 0~10 的水平上。这将呈现正偏斜,意味着分布的极值在量表右侧 0~10 的连续统上。

极端偏斜的分布迫使社会科学家在解释统计分析时必须谨慎,尤其是普通最小二乘回归的结果。因为,普通最小二乘回归的一项假设是自变量【Independent Variable】同因变量【Dependent Variable】有线性关系。另一假设是回归的误差项呈正态分布;在大样本(N>30)且因变量正态分布时,该假设自然成立。偏斜分布的因变量可能导致违反正态性假设。偏斜分布的自变量则难以产生最佳线性无偏估计【Best Linear Unbiased Estimation, BLUE】,至少当样本量较小时如此。

如下统计量可用以测定任意变量的偏斜度

$$y = \frac{\sum \left(\frac{y_i - \bar{y}}{s_y}\right)^3}{N}$$

其中,$\sum$ 为求和符号,s 为标准离差,N 为变量的观测数量。变量分布越接近正态,偏斜度统计量越趋近于 0。

偏斜分布是可以补救的。标准的纠正技术涉及变量的算术变换。最常用的变换

是对数【Logarithm】变换。其效果是破除变量值在少量数值周围的聚集,增加变量的散布。其他的变换包括抛物线变换,即求该变量的平方,或双曲线变换,其算法是 $1/y$。然而,在某些情况下,不存在实际的补救方案,研究者将不得不接受非正态分布的变量。

——Ross E. Burkhart

(赵锋译校)

参考文献

Weisberg, H. F. (1992). *Central tendency and variability* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-083). Newbury Park, CA: Sage.

偏斜(Skewness)

参见峰态【Kurtosis】。

斜率(Slope)

一条直线方程是 $Y=A+BX$,其中 Y 是自变量,A 是 y 的截距,B 是斜率,而 X 是自变量。斜率的计算可通过选取直线上的任意两点,再用两点间的纵向变化除以横向变化,其中横向变化是在 X 轴上两点间的增量,纵向变化是在 Y 轴上两点间的增量(或减量)(见图 1)。两个变量之间的正向关系表现为一个正的斜率,而负向关系表现为一个负的斜率。

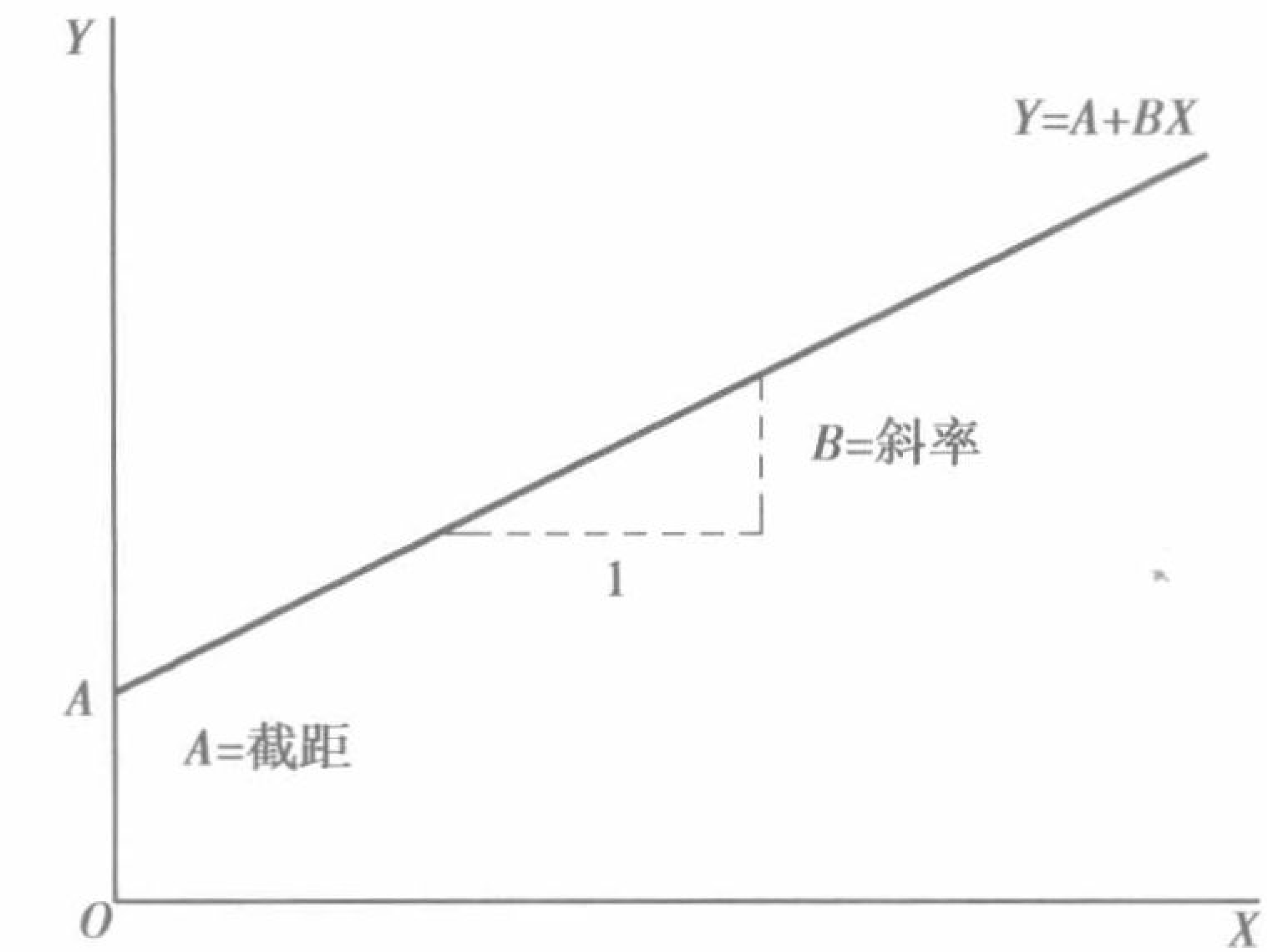


图 1 描绘了一条直线的斜率的图

当我们在社会科学研究中运用回归分析【Regression Analysis】时,数据从不会完美地拟合一条直线,而这就意味着预测存在误差。常见的简单回归方程是 $Y=A+BX+E$,其中 E 代表用给定的变量 X 预测变量 Y 时产生的误差。斜率则可解释成,在自变量上的一单位的变化伴随因变量变化的平均速率。在简单回归中(即方程只有一个自变量),斜率能够告诉我们收入和教育间的边际关系。举个例子,假设我们做收入(以美元为单位)对教育(以受教育年限为单位)的回归,并得到了一个值为 5 000 的斜率。如此一来,我们可以主张教育上一年的增量将伴随着平均 5 000 美元收入的增加。

在多元回归分析【Multiple Regression Analysis】中(即方程中有多个自变量),单个的斜率估计了每个自变量的偏相关关系。换言之,每个斜率反映了在控制模型中其他所有自变量后,某一单个自变量对因变量的效应。例如,假设我们做收入(还是以美元为单位)对教育年限和性别的回归,并得到教育的斜率为 4 000。这一斜率就可以解释为:保持性别不变,伴随着受教育年限每增加一年收入将平均增加 4 000 美元。

当两个定量变量之间的函数形式是线性时,通常报告斜率系数就可以了,因为它很容易解释。但是如果两个变量之间的函数形式是非线性的(即在自变量的取值范围内,回归线的斜率不是常量),那么一个单一的斜率系数就不能综括两个变量的关系。在这种情形下,就需要有几个斜率的估计值,通常还需要图表来辅助对关系的完全理解。例如,在多项式回归中(参见多项式方程【Polynomial Equation】),自变量有比回归线的方向变化的数量多 1 的斜率系数。在非参数回归中(如局部回归【Local Regression】),不存在斜率系数,而变量间的关系只能展示在一张图上。

斜率的概念也可以推广到线性模型的一般领域,如广义线性模型【Generalized Linear Models】和多层次分析【Multilevel Analysis】。在非线性模型中,如 Logistic 回归【Logistic Regression】中,通常来说,斜率最好通过计算拟合的概率而予以解释。

——Robert Andersen

(马妍译 赵锋校)

参考文献

Allison, P. D. (1999). *Multiple regression: A primer*. Thousand Oaks, CA: Pine Forge.

Lewis-Beck, M. S. (1989). *Applied regression: An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-022). Newbury Park, CA: Sage.

平滑(Smoothing)

平滑大致可以视为一种从数据中综括或提取重要特征的工具。在数据分析中,这一有效工具有很多应用。相对于条块的直方图【Histogram】,可以仅仅用平滑技术创建一个能够揭示一个连续【Continuous】变量的更多分布细节的直方图。在非参数【Nonparametric】估计中,同一平滑过程还常用于创建分布【Distribution】的函数【Function】形式。在考虑双变量【Bivariate】关系时,平滑技术还是一个有价值的工具,用以辨识非线性【Nonlinearity】情况下关系的函数形式(参见局部回归【Local Regression】)。在单一变量的(单变量【Univariate】)时间序列分析中,在解释性研究中,为了综括变量中的长期变化,可以运用平滑技术。对一个时间序列的未来值的预测【Forecasting】,如通货膨胀或增长率,平滑技术也特别有用。如果数据点有一个不规则的分布,无论是归之于历时测量的一个变量的白噪声【White Noise】部分,即一个变量的密度函数上的突然变化,还是一个非线性的双变量关系,平滑技术通过取出数据的系统变化或模式,可以提供有用的信息。

直方图平滑

相较于条块的使用,一个直方图的平滑能更好地揭示一个连续变量分布的特性。条块的缺点在于,由于条块宽度和条块原点选择的随意性,条块直方图会掩盖变量分布的连续性,进而可能错误地呈现数据。作为一种常规的分析实践,平滑容许更好地检查分布,在使用专门的估计量【Estimator】时,还容许更好地评估正态性【Normal】假定。

设若有一个非正态分布的连续数据,直方图平滑还常用在非参数估计中,以创建该

分布的函数形式。当有理由认为参数估计量的假定未被满足时，如数据是非正态分布的，常会使用非参数估计。当变量的潜在分布缺乏理论设定【Specification】时，样本数据可提供分布的函数形式。

对一个变量分布的常规分析，以及在非参数估计中，为了探测分布的函数形式，一项平滑技术可以获得分布形态的最佳表现。就一个变量 X 而言，平滑技术本质上使用接近于变量的特定值 x_i 的若干观测点来构造直方图的密度。平滑的程度同平滑参数（也称为“带宽”）的量级正相关。平滑参数即观测值相对平滑操作所施用的值的比率。一个相对小的平滑参数值会导致较为不平滑，因为相较于离变量值 x_i 较远的观测值，越接近于值 x_i 的观测值相对地被给予更大的权数。相反地，一个相对大的平滑参数值会带来更佳的平滑，因为给予临近观测值的权数相对于更远的观测值趋于一致。应注意避免“平滑不足”或“过度平滑”，因为两者的结果都是引入了偏倚。若数据中保留了过多的噪声，以至于直方图不能揭示分布的有意义的特征时，即出现平滑不足，也称为“过度拟合”。过度平滑或“拟合不足”指一个掩盖了数据重要特征的直方图。为确保选出的分布对平滑参数的量级上小变化不是非常敏感，一个有效的实践是用各种平滑参数进行试验。

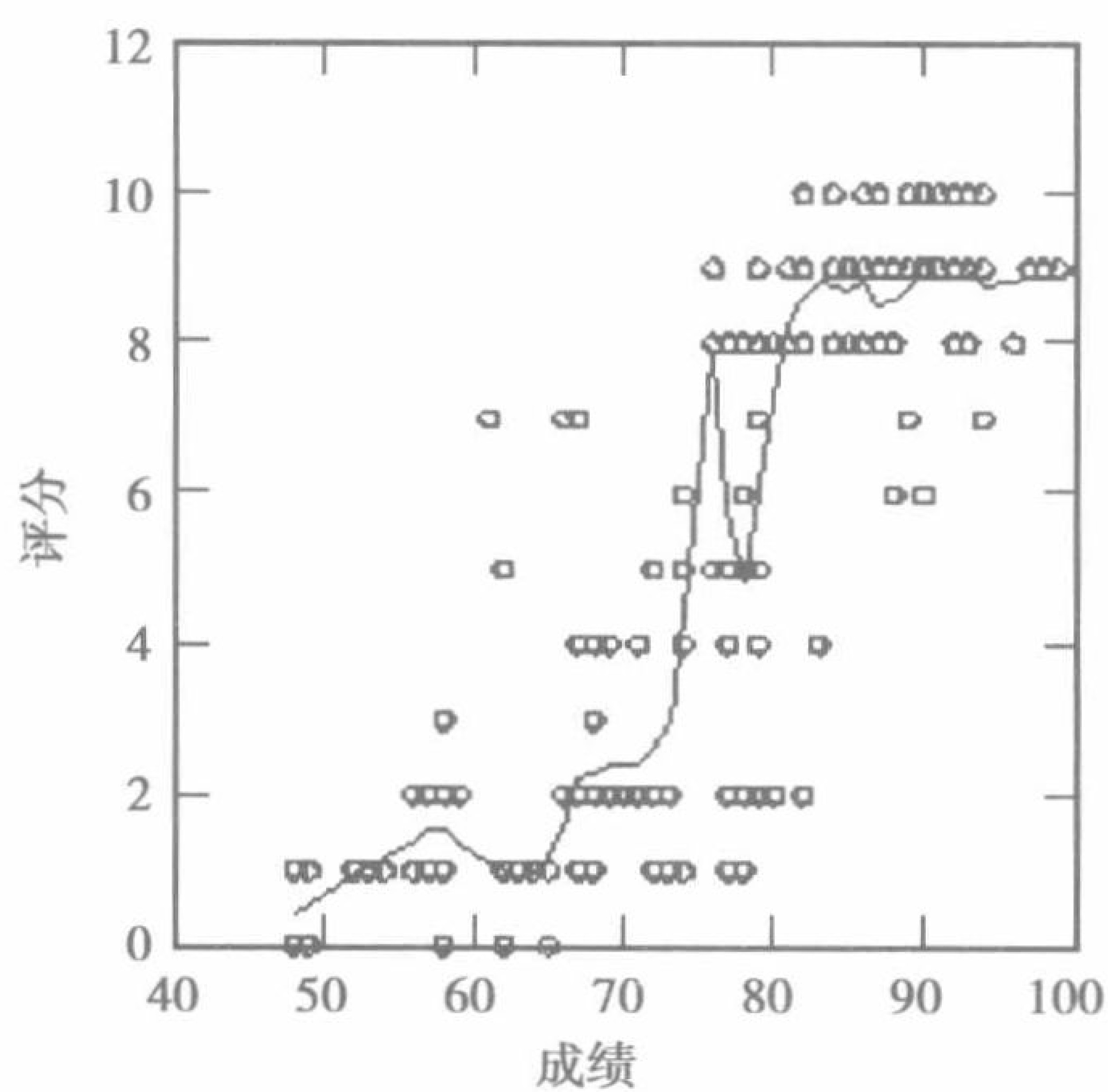
双变量或散点图平滑

在探查双变量关系的函数形式时，平滑也很有用。通常，可以绘制一个散点图，来探查双变量关系（线性的【Linear】、曲线的【Curvilinearity】等）的函数形式。遗憾的是，函数形式经常不清晰。平滑诸系列的数据点可以综括“变量 Y 的分布的集中趋势如何在变量 X 的分布不同位置上变化”（Jacoby, 1997, p. 63）。就如同在非参数估计中探查单变量分布的函数形式，平滑一个双变量分布的目的是允许数据指明关系的函数形式。有关常见的双变量平滑方法的说明，参见词条局部回归【Local Regression】。再次注意，要避免过度平滑或平滑不足。雅各比（Jacoby，1997）提出了一些有用的策略来避免这些错误。例如，密度线没有穿过一组观测点或者在平滑后的残差中留有结构都表示过度平滑（或拟合不足）。图 1 分别给出了双变量分布的平滑不足、过度平滑和恰当平滑的例子（即一个假设的班级，X = 以百分比计算的成绩，Y = 学生对教授的评分）。

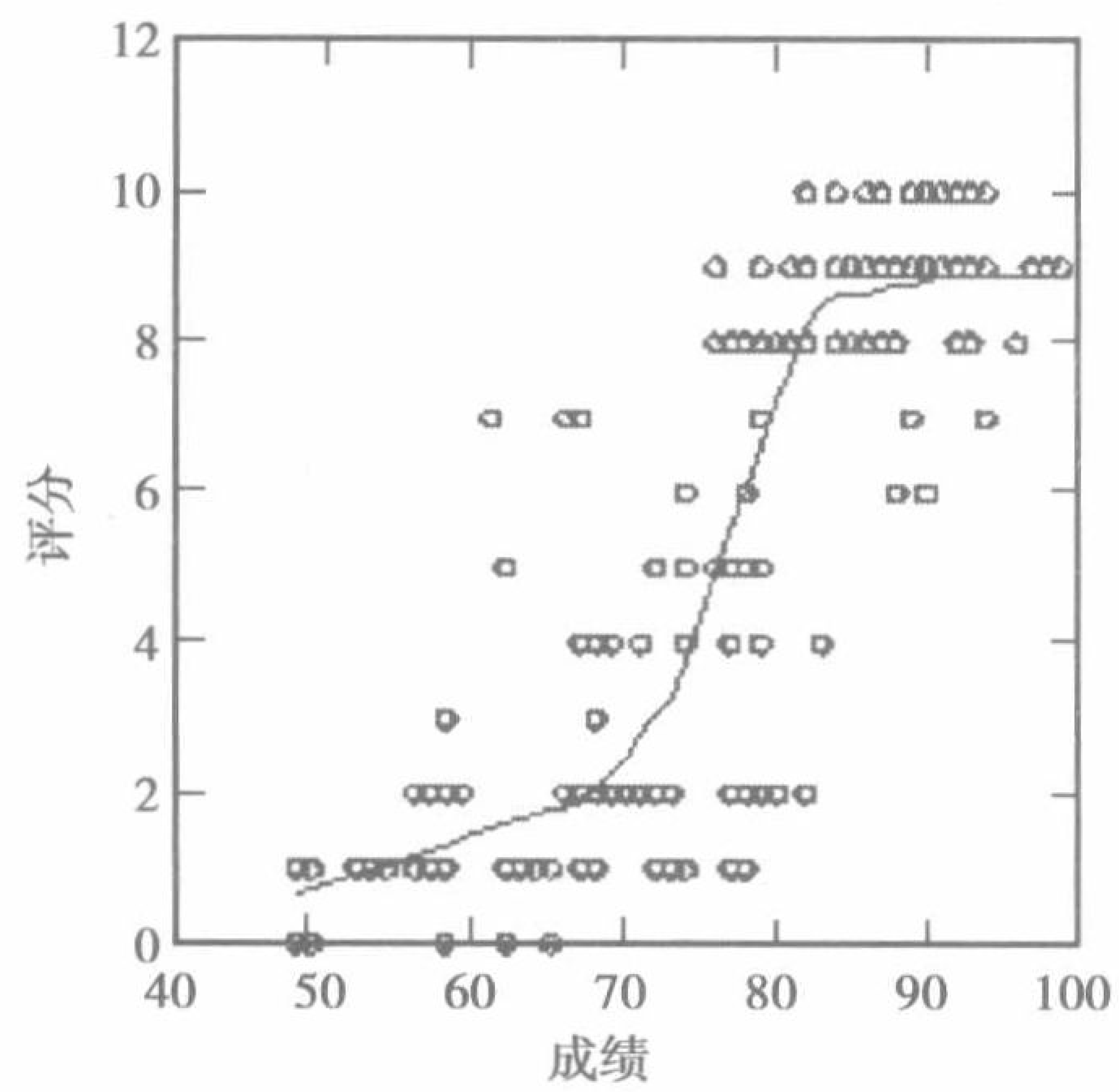
时间序列平滑

当平滑一个时间序列的变量时，数据本质上被视为由真实的潜在模式部分加残差【Residual】噪声构成，或者由“平滑的”部分加“毛糙的”部分构成。与一般平滑类似，时间序列平滑的目标是：既不是过度平滑的（由于过度的平滑，失去了时间序列的潜在模式的变化方式），也不是平滑不足的（一个显示了过量噪声的时间序列，以至于无法从噪声中识别其潜在模式）取出时间序列的真正潜在模式。在两种情况下，潜在模式都不容易识别。在时间序列预测中，过度平滑或平滑不足会显现为一个较大的预测误差。

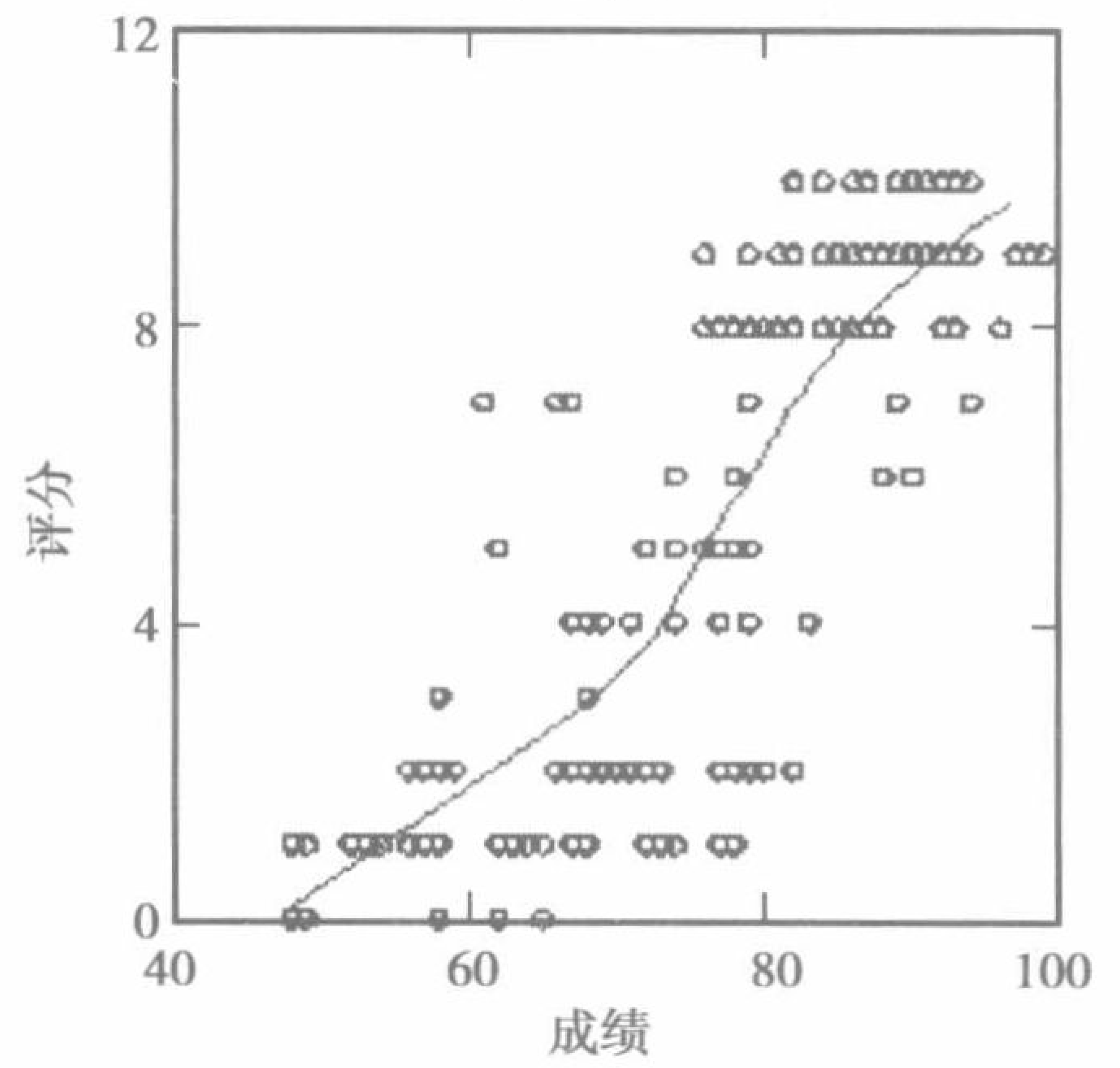
就单一变量的时间序列数据，存在着几种不同的平滑技术（Jacoby, 1997; Mills, 1990）。在某种程度上，所需的技术取决于分析的目的。例如，指数平滑（exponential smoothing）可能适合于预测分析，因为这种技术赋予较近的观测值的权数远大于较远



(a)平滑不足



(b)平滑



(c)过度平滑

图 1　平滑示例

的观测值。就探索的目的而言,一个简单的方法,如移动中位数(running medians)或移动加权平均(running weighted averages)就足以满足要求。本质上,这些技术通过用临近的多个数据点的中位数或加权平均值替代每一个数据点,从而平滑时间序列。用于校准每个观测值的最小的序列,或临近数据点的数量是 3。例如,为确定数据点 x_t 的平滑值,使用序列 3 的加权平均公式如下:

$$s_t = (w_i)x_{t-1} + (w_i)x_t + (w_i)x_{t+1}$$

其中,s_t 是平滑值,而 w_i 是分配给每个观测值的权数(围绕 x_t 对称)。然后对每一个 x 值重复运用这个公式。部分地,恰当的序列取决于所需的平滑程度。

幅度较大的平滑也可能通过重复迭代同一项平滑技术或者顺序运用不同的平滑技术实现。当然,阶数越大且(或者)平滑处理的次数越多,过度平滑的敏感度就越高。

——Jacque L. Amoureux

(马妍译　赵锋校)

参考文献

Jacoby, W. J. (1997). *Statistical graphics for univariate and bivariate data*. Thousand Oaks, CA: Sage.

Kennedy, P. (1998). *A guide to econometrics* (4th ed.). Cambridge: MIT Press.

Mills, T. C. (1990). *Time series techniques for economists*. Cambridge, UK: Cambridge University Press.

滚雪球抽样(Snowball Sampling)

滚雪球抽样可以被定义为这样一种收集研究对象的技术,即通过确认一个最初的研究对象,再请他提供其他参与者的名单,进而收集整个研究对象。这些人自己可以为扩展联络和调查的网络提供可能性。这一策略曾经主要用来克服隐蔽总体的理解和抽样问题,如离轨者和社会孤立人群(Faugier & Sargeant, 1997)。滚雪球抽样可以放在这样一组更广泛的方法论的范围内,即充分利用已有回答者的社会网络,为研究者提供不断增加的潜在联系群体。

滚雪球抽样有时是对这样一种技术的一个误称,而该技术在传统上更多地与定性研究【Qualitative Research】联系在一起,并作为接近隐藏总体的一种权益策略。由于与传统的随机选样和代表性相关的许多假定相矛盾,这一方法在社会科学领域中遭遇到形象问题。然而,这一技术有利于接触到那些被剥夺的、被社会污名化的,以及精英人群。桑德斯在他的关于伦敦南部的政治权力和影响的研究中,使用了一种声望法,即让他的联系人给他指出那些在这一地区被视作拥有权力的人(Saunders, 1979)。通过这种方法,就可以得到各种相当重要,但又是潜在地自我包含的社会交往系统,而这些社会交往系统还能够让我们洞见到政治权力的知觉基础。

滚雪球抽样作为一种技术也有所发展,相关的文献显示,这一技术有朝着更加成熟的抽样框和误差估计发展的趋势,也显示出它正逐渐作为研究者工具箱中的一项被采纳的工具。肖等(Shaw, Bloor, Cormack,& Williamson,1996)用示例说明,他们使用滚雪球抽样来估计难以寻找到的群体的现状,例如无家可归者。现在,这一被视作有局限性的定性工具成为一项系统的调查方法,在社会科学研究领域获得了新的应用。然而,尽管选择性偏倚可能通过生成更大样本量和结果的重复【Replication】来解决,也还是有看门人偏倚的问题,即联系人群被阻止与研究者接触。然而,其他群体可能自身就由那些高度原子化且高度孤立的个体(他们的社会网络相对薄弱,与他人联系困难)构成。

滚雪球抽样可以用于两个主要目的:第一,作为一种非正式的方式找到一个目标总体,而这时,一项研究的目的可以是探索性的或是新奇的;第二,作为一种较正式的方法,对那些难以使用传统方法(如家庭抽样调查)抽取个体的总体作统计推论。这一技术早期的例子可在怀特(Whyte, 1955)的《街角社会》(*Street Corner Society*)中看到。在这本著作中,怀特用最初的联系人产生可以用于研究帮派动态的背景脉络和关系(尽管这包含了从参与观察【Participant Observation】到为了更近期的访谈研究而使用滚雪球抽样技术的方法转换)。

尽管一些人把滚雪球策略视为琐碎的或意义不明确的,但是这一方法的主要价值在于能够克服获取回答人的困难,当调查的回答人或者数量很少,或者由于信任的问题难以取得初始联络时。在这些情况下,这类"传递链"技术可以让研究者渗透到内部或成为团体成员,从而进入用那些传统方法很难进入的环境。

——Rowland Atkinson
John Flint
(王玥译 赵锋校)

参考文献

Faugier, J., & Sargeant, M. (1997). Sampling hard to reach populations. *Journal of Advanced Nursing*, *26*, 790-797.

Saunders, P. (1979). *Urban politics: A sociological interpretation*. London: Hutchinson.

Shaw, I., Bloor, M., Cormack, R., & Williamson, H. (1996). Estimating the prevalence of hard-to-reach populations: The illustration of mark-recapture methods in the study of homelessness. *Social Policy and Administration*, *30*(1), 69-85.

Whyte, W. F. (1955). *Street corner society: The social structure of an Italian slum* (2nd ed.). Chicago: University of Chicago Press.

社会期望偏倚(Social Desirability Bias)

社会期望偏倚是,当人们报告有关自己的情况时,根据一种可为社会接受的方向,对问题予以回答的倾向。这一回答偏倚【Response Bias】主要发生在那些涉及个人的或社会的敏感内容的题项或问题上。社会期望被认为是一个随个体而变化的人格变量(Crowne & Marlowe, 1964),同时,只有某些个人会表现出这种偏倚。克朗尼和马洛认为,社会期望是个体对赞同的需要的一个指标,而这一需要会导致一项回答偏倚,但是这一需要本身不是一项回答偏倚。

人们最常用克朗尼和马洛(Crowne & Marlowe,1964)的社会期望量表估计作为一个人格变量的社会期望。该量表由关于自我陈述的一组是/否问题组成,其中每个问题反映出的事项,或者是期望的,或者是不期望的(如"我总是在投票前核实候选人的资格")。由于极端的措辞"我总是"及"我从来不",所以,就个人而言,在一组题项中,几乎不可能有很多的陈述在字面上是真的。高分反映出个人的企图,即无论有意地,抑或无意地,将自己置于一个受到认可的光照下,同时让自己表现得可以为社会所接受。也有其他的量表可用于估计社会期望,例如,保卢斯(Paulhus,1984)量表可以估计两个维度的社会期望:欺他和自欺。

尽管有普遍的担忧,然而没有很多证据表明社会期望是那些有赖于自我报告的研究中普遍存在的问题。例如,穆尔曼和波德萨可夫(Moorman & Podsakoff,1992)研究了组织研究中的社会期望的普遍程度和效应。他们的结论指出,社会期望同被考察的变量间的关联显得相当小,而且同其他变量相比,社会期望对结果只有很少的偏倚效应。然而,也存在一些研究领域,其中社会期望是一项潜在的问题,如心理调试领域。这一领域关注的是内在的心理学状态(如焦虑和自我评价),而这些状态被认为是社会期望的,同时,最好能通过自我报告而予以估计。

有两种方法可以处理社会期望偏倚。一种是通过设计工具和方法来减少它对估计的影响。这种方法可以通过精心选择那些同社会期望没有关联的量表题项来实现。不过,这种方法只适用于某些概念,而非所有感兴趣的概念。第二种是估计社会期望独立于其他变量的效应,并控制它。通过把那些在社会期望量表上获得高分的个体排除于分析之外,或通过统计控制【Statistical Control】(求偏相关系数),将社会期望的效应排除于其他变量之外,即可以实现。然而,这种方法可能导致方法本身带来的偏倚效应。因为人们不能确定,社会期望同其他变量的关系是否能完全归之于偏倚,还是反映了需要赞同的人格变量同其他兴趣变量之间的真实关系。

——Paul E. Spector

(王玥译 赵锋校)

参考文献

Crowne, D., & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence*. New York: Wiley.

Moorman, R. H., & Podsakoff, P. M. (1992). A meta-analytic review and empirical test of the potential confounding effects of social desirability response sets in organizational behaviour research. *Journal of Occupational and Organizational Psychology*, *65*, 131-149.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*, 598-609.

社会关系模型(Social Relations Model)

社会关系模型(Kenny & La Voie, 1984)用于分析来源于两人之间的互动或者小组成员之间的互相评分的成对的(dyadic)或者关系的数据。不同于社会测量【Sociometry】研究,社会关系模型的测量层次【Level of Measurement】是定距的(如7点刻度),且是非类别的(如是或否)。一般而言,数据以人群作为收集对象,但是也可以有其他收集单元(如机构)。该模型可被视为一种复杂的多层次【Multilevel】模型。

最常见的社会关系设计是循环研究设计。在这种研究设计中,小组中的每个人都与其他个人互动,并对每一个他人评分,而数据则收集自每个成员对的两个成员。通常进行的是多组循环。其他设计也有可能。最低限度是每个人都对多个他人进行评价或者与之互动,并且每个人都被其他多个人评价或与之互动。

社会关系模型有三种主要的效应类型。行动者效应是指在一组不同的同伴面前,一种给定的行为的个人平均水平。例如,汤姆在信任变量上的行动者效应测量了他趋向于信任他人的一般程度。同伴效应是指一个人从一组不同的同伴那里引出的某种反应的平均水平。汤姆的同伴效应测量了其他人倾向于信任他的平均程度。关系效应是指一个人对其他某个特定个体,高过和超出行动者效应和同伴效应之上的行为。例如,汤姆对玛丽在信任变量上的关系效应测量了在控制汤姆信任他人的一般倾向和玛丽被其他人信任的一般倾向后,汤姆对玛丽的信任。关系效应是有方向性的或非对称的,比如,汤姆对玛丽的信任可能多于或者少于玛丽对他的信任。为了将关系效应从误差方差中区别出来,就需要有所测概念的多个指标,或者是历时的,或者使用不同的量度工具。

社会关系模型的关注点不在于估计特定人和特定关系的效应,而在于估计各类效应的方差。因此,就一项关于聪明人如何看待彼此的研究而言,其兴趣点在于是否存在着行动者、同伴和关系方差。行动者方差评估人们是否视他人为具有相似的智力水平;同伴方差评估在进行对他人的智力评分过程中,人们之间是否同意彼此;关系方差评估在多大程度上对智力的知觉与配对(dyad)唯一相关。

社会关系模型也能用于相互作用(reciprocity)的具体研究。例如,可以用该模型研究那些倾向于喜欢他人的行动者是否也会被他人喜欢(一般的相互作用),以及是否存在着关系效应的相互作用(成对的

相互作用）。

已经有大量的研究使用该模型。最集中的应用则出现在人际知觉领域（Kenny，1994）。同时，就揭露的相互作用、小群体过程，以及裁判准确性中的个体差异等议题都有该模型的应用。

——David A. Kenny

（马妍译　赵锋校）

参考文献

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis.* New York: Guilford.

Kenny, D. A., & La Voie, L. (1984). The social relations model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 18, pp. 142-182). Orlando, FL: Academic Press.

交往图（Sociogram）

一张交往图即通过一张图呈现个体（或者任意一个可能的分析单位）间的关系。交往图有助于映射出关系的结构。它们通常是网络分析【Network Analysis】或社会测量【Sociometry】体系下的研究结果。

——Alan Bryman

（马妍译　赵锋校）

社会测量（Sociometry）

这一术语偶尔用来指社会科学中与通过量表【Scale】进行测量有关的测量方式。它更常见的用法涉及一种强调绘制分析单位（个体、群组、组织等）间的关系图的研究方法。越来越多地，社会测量被称作网络分析【Network Analysis】，而且大多数原来被称作社会测量的研究也已经被纳入这一领域。

——Alan Bryman

（马妍译　赵锋校）

软系统分析（Soft Systems Analysis）

软系统分析（SSA）是一种在复杂系统中调查问题和计划变革的方法。它也可以用来设计新的系统。它最初是由兰卡斯特大学（Lancaster University）的彼得·切克兰德发展出来的（Checkland，1981）。这一方法最好被视为一个实践的工作工具，虽然它也广泛地用在应用研究中。一个核心的问题是相对于"硬"系统，"软"的意义是什么。软系统思维的核心理念是，人们看世界和解释世界的方式是不同的。个人在观点上的差异并非是无效的来源，也不是数据中的"噪声"；相反，区别反映了实在的本质。多元主义是规范。在复杂的系统中，个人或群体很可能对系统是如何运作的，系统出现了什么问题，以及应该怎样去改变系统，构建出非常不同的观点。

尽管认识到 SSA 可以有不同的使用方式，然而最简单的方式是用最直接的形式去描述（Naughton，1984）。这种方法通过一

系列相对正式和有较好结构的不同阶段组织起来，从而可以让使用者使用。在实践中，在不同阶段可能有一定的重复。分析者最初收集关于问题的境况（第 1 阶段），接着以图片式的方式（“丰富的画面”）呈现（第 2 阶段）。方法的使用者（即分析者和系统的参与者）明确地试图以多种不同的方式来审视系统，以期发现那些可以带来新意（一个“相关的系统”）的观点（第 3 阶段）。他们选择与问题境况有关的新的视角，发展出一个概念模型，即系统在逻辑上应该做些什么才能满足这些新观念的要求（第 4 阶段）。然后将这个模型与现存的问题境况作比较，看看是否有任何改变的经验教训（第 5 阶段）。参与者讨论这些经验教训，以决定哪些可以使用（第 6 阶段）。获得同意的变革于是被执行（第 7 阶段）。如果新的观念（来自第 3 阶段）看起来无法帮助参与者，再尝试另外一种视角（返回到第 3 阶段）。

用一个例子可以帮助说明这个方法的重要组成部分。西蒙和克莱格（Symon & Clegg，1991）在英国一个大的航天公司研究一种计算机辅助设计系统和计算机辅助生产系统（Computer Aided Design Computer Aided Manufacturing，CADCAM）的实施。这个研究是在 1988—1990 年开展的，在 1996 年又开展了追踪研究。经过了广泛的资料收集阶段（第 1 阶段和第 2 阶段），研究者认为公司对实施过程采取了不适当的（并且是很局限的）观点（或相关系统），同时还提供了一个不同的观点（第 3 阶段）。旧的相关系统把实施过程看作一个简单的技术实施项目，而新的系统则强调组织变革。新的概念模型就产生了，其中一项要求满足 CADCAM 实施的新观念的需要（第 4 阶段）。新的模型强调，资源、计划、参与、教育，以及支持和评估的诸项需要。这些需要与实际的实施活动相匹配，可以帮助发现实践中的一些重要的缺口。这样就带动产生一系列的行动建议（第 5 阶段）。经过讨论（第 6 阶段），一些建议被公司采纳并执行（第 7 阶段）。

研究者继续他们的工作，并在 CADCAM 已经有非常广泛的使用时，进一步评估它产生的影响，再次将概念模型与实践进行匹配（回到第 4 阶段）。进一步的缺口被识别出来，并给公司提出第二轮的行动建议（第 5 阶段）。其中一些被再次认可和执行（第 6 和第 7 阶段）。

随后 6 年的追踪研究揭示出 CADCAM 的执行获得了成功，因为技术已经有全面的实施，广泛的使用，且很受欢迎。其中大多数初始目标都已经得到实现。研究者发起的不同行动（例如组成 CADCAM 用户小组）持续发挥功能。然而，CADCAM 实施还没有达到设计和生产之间想要追求的整合水平，主要是因为并没有尝试去实现这两个强大的功能之间整合所必需的结构变革。

从这个例子可以看到，SSA 与行动研究【Action Research】有着明显的相似性。在这个例子中，我们可以把 SSA 看作一种特别的行动研究。当然，行动研究还包括不同的方法（如调查反馈）。而且，一些 SSA 改变项目更加关心的是行动而不是研究，因此严格地讲不应该把 SSA 标为行动研究的范例。

SSA 有一些基本特征（Clegg & Walsh，1998）。SSA 是高度参与性的，同时，它提供了在一个复杂系统中分析问题和计划变革过程的相关结构和组织框架。在这个过程的不同阶段，它积极促进创新思维，要求仔细的分析和逻辑的建模。随着它的应用，该方法可以被内化，同时，在我们看来，对于参与研究和应用的技术人员来说，它可以成为一个非常有效的附加工具。

——Chris Clegg
Susan Walsh
（王玥译　赵锋校）

参考文献

Checkland, P. B. (1981). *Systems thinking, systems practice*. Chichester, UK: Wiley.

Clegg, C. W., & Walsh, S. (1998). Soft systems analysis. In G. Symon & C. M. Cassell (Eds.), *Qualitative methods and analysis in organisational research: A practical guide*. London: Sage.

Naughton, J. (1984). *Soft systems analysis: An introductory guide*. Milton Keynes, UK: Open University Press.

Symon, G. J., & Clegg, C. W. (1991). Technology-led change: A study of the implementation of CADCAM. *Journal of Occupational Psychology*, *64*, 273-290.

所罗门四组实验(Solomon Four-Group Design)

所罗门四组实验(Solomon, 1949)是一种用于评估前测敏感性【Pretest Sensitization】效应——即是否仅仅由于实施一项前测就影响了后续测验得分——的合理(plausibility)方法。例如,在一项词汇实验中,接受了前测的参与者可能会决定通过查字典来掌握前测中出现的一些不熟悉的单词。在后测中,他们在词汇测验中的得分将会高于他们没有接受前测情况下的得分。元分析【Meta-Analysis】的结果表明,前测敏感性确实存在,尽管相比于其他的测量,该效应更普遍地出现于某些类型的测量中;同时,前测敏感性的出现依赖于前测和后测间的时间跨度(Willson & Putnam, 1982)。

在所罗门四组设计中,研究者随机地将参与者分配入四个组别之一,而这四个组别则根据两个完全交叉的因素进行构建——处理(如处理【Treatment】或控制组【Control Group】)和是否实施前测。因此,四个组分别是:(1)有前测的处理组;(2)无前测的处理组;(3)有前测的控制组;(4)无前测的控制组。随后的分析要检验:处理组的效果是否比控制组好;接受了前测的组的得分是否不同于没有接受前测的组的得分;以及是否存在前测与处理的交互效应。

所罗门四组设计同时提供了与内在效度【Internal Validity】和外部效度【External Validity】相关的信息。例如,考虑到内在效度,许多单组准实验【Quasi-Experiment】同时使用前测和后测。如果出现了前测敏感性效应,那么对处理效应的判断就可能是错误的,因为即使处理没有任何效应,从前测到后测的得分也会出现变化。然而,在随机化的实验【Experiment】中,前测敏感性效应不会影响内在效度,因为它们在不同的条件上是随机分布的;但是如果效应的大小取决于是否进行了前测,那么前测敏感性效应就可能同外在效度相关。

计算机文献搜索发现有102项研究使用了该方法或者该方法的修订样式,其中有超过半数(55项)的研究是在过去10年以内发表的。其中许多研究发现不存在前测效应(如 Haight, Michel, & Hendrix, 1998)。杜克斯、厄尔曼和斯坦(Dukes, Ullman, & Stein, 1995)在他们所使用的四个因变量【Dependent Variable】中的一个因变量(对同辈压力的抵抗)上发现了前测效应,并由此建议在比较组有前测和无前测的情况下,处理效应潜在地缺乏概推性。

——William R. Shadish

M. H. Clark

(马妍译　赵锋校)

参考文献

American Psychological Association. (2002, October 28). *PsycINFO database records* [electronic database]. Washington, DC: Author.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: RandMcNally.

Dukes, R. L., Ullman, J. B., & Stein, J. A. (1995). An evaluation of D. A. R. E. (Drug Abuse Resistance Education), using a Solomon four-group design with latent variables. *Evaluation Review*, *19*(4), 409-435.

Haight, B. K., Michel, Y., & Hendrix, S. (1998). Life review: Preventing despair in newly relocated nursing home residents: Short-and long-term effects. *International Journal of Aging and Human Development*, *47*(2), 119-142.

Michel, Y., & Haight, B. K. (1996). Using the Solomon four design. *Nursing Research*, *45*(6), 367-369.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Solomon, R. L. (1949). An extension of control group design. *Psychological Bulletin*, *46*, 137-150.

Walton Braver, M., & Braver, S. L. (1996). Statistical treatment of the Solomon four-group design: A meta-analytic approach. *Psychological Bulletin*, *104*(1), 150-154.

Willson, V. L., & Putnam, R. R. (1982). A meta-analysis of pretest sensitization effects in experimental design. *American Educational Research Journal*, *19*, 249-258.

分拣(Sorting)

在分拣方法中,回答人被呈予一组对象,并被要求将对象分入一些群组中。作为一种数据收集方法,分拣假定研究者已经挑选出一个特定的范围或领域,并且以一张示例清单(由一个单词、短语、照片、图片、题项,或现实的物构成)对这范围或领域进行了界定。这些对象通常已经事先由某些程序挑选出来,诸如自由列表或分类学构建。通常用一张卡片,向回答人呈现这些对象,再要求回答人将它们分入不同组或堆(即成堆分拣),以便相似的对象被放入相同的组。于是,这些分组代表了领域内独特的类别或种类,并且共同组成一个定类尺度。

分拣方法尤其常用于语言学研究(欲研究类别形成的过程),以及社会和文化世界的知觉研究(为了考察参与者使用哪些意义类别和意义划分)。分拣法既具有能处理数量众多的对象的优点,同时也是回答人表示乐于进行的一项任务。

基本分拣中的变化样式

在基本形式(自由分拣)中,可能形成任意的组数和任意的规模。对分组数量和种类的限定界定了分拣的变化样式。

- 当存在预先设定的群组数量时,是确定分拣。
- 当群组的数量,以及群组之间的等级序列都固定时,是分级分拣,好似在一张李克特量表【Likert Scale】上。
- 除了有固定和序列的群组,还要求在所有群组中的对象服从一个特定的正态分布,这是 Q-分类。

一旦形成了群组,即邀请回答人命名和

(或)描述它们,从而就类别的语义给出有价值的解释数据。

更复杂的分拣类型包括如下：

• 多重(multiple)或连续分拣(successive sortings)(Boster, 1994)。在这种方法中,回答人根据同样的或不同的准则对对象进行若干次分拣。

• 细化(finer)/粗化(coarser)分拣(或称为群组、分割、附加分拣, GPA: Group-, Partition-, Additive-sorting)(Bimler & Kirkland, 1999)。这种分拣出现在,当最终的分拣类别(a)被合并在一起,从而形成一个更加粗略的分拣,以及(b)被进一步分解,从而形成一个具有更细致的分类的类别。一种流行的分拣形式是汇聚层级分拣(agglomerative hierarchical sorting)(或层级聚类法, hierarchical clustering method)(Coxon, 1999),在这种方法中,回答人开始将每个对象分入一个单一的群组("分离器"分割),再形成一系列连续的更为一般的分拣,直至所有的对象都进入一个唯一的群组("打包者"分割)。

• 重叠分拣(overlapping sorting)。在这种方法中,回答人可能被允许将同一对象划入不止一个群组中,但是在这种情形下,将不能使用传统的统计方法,因为数据将不再形成一种分割的状态。

分拣数据的分析

一旦完成数据收集,分拣数据就能根据对象的相似性而予以分析,或者使用共现量度(co-occurrence measures)(Burton, 1975)来衡量一组对象在回答人的群组中一并出现的频率;或者用最少移动分割量度,衡量分拣的相似性(Arabie & Boorman, 1973)。这些量度是用多维度量【Multidimensional Scaling】和层级聚类法来表现的,同时,就这类数据,塔卡宁(Takane, 1981)还发展出联合的和三维度量的模型。

——Anthony P. M. Coxon

(马妍译　赵锋校)

参考文献

Arabie, P., & Boorman, S. A. (1973). Multidimensional scaling of measures of distance between partitions. *Journal of Mathematical Psychology*, *10*, 148-203.

Bimler, D., & Kirkland, J. (1999). Capturing images in a net: Perceptual modeling of product descriptors using sorting data. *Marketing Bulletin*, *10*, 11-23.

Boster, J. S. (1994). The successive pile sort. *Cultural Anthropology Methods*, *6*(2), 7-8.

Burton, M. L. (1975). Dissimilarity measures for unconstrained sorting data. *Multivariate Behavioral Research*, *10*, 409-424.

Coxon, A. P. M. (1999). *Sorting data: Collection and analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-127). Thousand Oaks, CA: Sage.

Takane, Y. (1981). Multidimensional scaling of sorting data. In Y. P. Chaubey & T. D. Dwivedi (Eds.), *Topics in applied statistics*. Montreal: Concordia University Press.

稀疏表(Sparse Table)

稀疏表是一种由两个或多个离散变量构成的观测值的交叉分类表,且表中有许

多单元格仅有很小或为零的频次。稀疏的列联表【Contingency Table】经常出现在观测值的总数相对于单元格的数量较小的时候。例如,考虑有一个张表格含有 $N=4$ 个单元格,且有 $n=12$ 个观测值。如果观测值均匀地分布在四个单元格里,那么每个格子的最大可能频次就较小(即,$n/N=3$)。更进一步,如果在四个单元格中的一个单元格中观测值的出现是一个罕见事件,那么要在这个单元格中得到观测值就需要一个很大的样本。再考虑第二个例子,有一个四个变量构成的交叉分类表,每个变量有 7 个类别,因此表格有 $N=7^4=2\,401$个单元格。要保证所有的单元格都有非零的频次,就需要有远大于 2 401 的样本量,同时,要保证所有单元格内的频次都大得足以应用统计检验和统计模型,就需要有更大更多的样本量。

单元格稀疏会使标准的统计假设检验变得无效,例如独立性或模型拟合优度的卡方检验。对于一个卡方分布而言,比较检验统计量[如皮尔逊卡方统计量(Pearson's Chi-Square statistic)或似然比统计量(likelihood ratio statistic)]的合理理由主要依赖于有"大的"样本,而"大"意味着就每个单元格而言,有大于或等于 5 的期望值(频次)。在缺乏大样本的情况下,检验统计量欲与之比较的概率分布是未知的。解决这个问题可能的方法包括:使用一个替代统计量进行精确检验,或者运用再抽样或蒙特卡洛方法来趋近检验统计量的概率分布。

稀疏表经常包括单元格为零的频次,而这可能导致估计问题,包括偏倚的描述统计量(如优势比【Odds Ratio】)、对数线性模型【Log-Linear Model】参数的估计,以及用于拟合模型的计算算法的难题。一个估计问题是否存在依赖于数据中非零频次的模式和所要估计的特定模型。当表格的边缘上存在零频次的单元时,参数是无法估出的。例如,表 1 包含了 1996 年综合社会调查中由以下两个问题及其答案选项构成的交叉分类表:(a)"据你所知,你认为基因检测是利大于弊吗?"答案选项是"利大于弊""弊大于利"和"视情况而定";(b)"关于基因筛查你认为自己听到或看到了多少?"答案选项是"很多""不太多""一无所知"。就第一个问题,没有人回答"视情况而定"。就独立对数线性模型而言,"视情况而定"的边缘效应的参数无法进行估计。估计这个参数所需要的信息是没有观测值的列边缘值。

表 1　带有零频次的二维表示例

		基因检验:利弊如何?			合计
		利大于弊	弊大于利	视情况而定	
对基因检验了解多少?	很多	54	10	0	64
	不太多	170	88	0	258
	一无所知	17	14	0	31
合计		241	112	0	353

在较大的数据集中,要侦测一个估计问题则更加困难。存在一个估计问题的标志是,标准误的极大的估计值,或者,在计算机输出上出现类如"收敛失败"的错误消息。解决估计问题的方法包括:在模型中增加参数,使得导致问题产生的空单元格恰好被拟合;给表中所有单元格加上一个非常小的值(如 1×10^{-8});或者使用另一种备选的估计方法。当空单元格的存在是因为它们在理

论上本不可能存在时(如一个男性有月经问题),那么应当在对数线性模型中为每一个空单元格增加一个额外的参数。增加的参数会将这些空单元格从分析中排除。

——Carolyn J. Anderson

(马妍译 赵锋校)

参考文献

Agresti, A. (1996). *An introduction to categorical data analysis*. New York: Wiley.

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley.

空间回归(Spatial Regression)

空间回归分析由一些具体的方法构成,其目的在于把握由于将回归方法用于空间(地理)数据或用回归方法去模型空间交互作用所引起的复杂状况。存在两个涵盖广泛的空间效应类别。一个被称作空间异质性【Heterogeneity】。当在整个观测值上,模型的参数(回归系数、误差项方差)、函数形式自身,或两者同时都不稳定时,它即出现。这是结构不稳定的一个特例,可以用常用的方法(模型异质性、改变系数、随机化系数)予以处理。当变化呈现出一种空间结构时,这就成为空间异质性,而利用这种空间结构能更有效地建模这类异质性。例如,在暴力犯罪的决定因素分析中,发现美国南部州有着不同于其他州的回归系数。由于这些州具有确定的地理位置,这就被称作一个空间政体,即空间异质性的一个具体形式。

空间效应的另一种形式被称作空间依赖(spatial dependence)(或空间自相关【Autocorrelation】)。当回归模型中的误差项是空间相关的,或当空间交互作用作为解释变量被纳入回归设定中时,这种效应即出现。空间误差项自回归是带有非零对角元素的误差项协方差矩阵的一个特例。当非零元素的模型遵循一个空间交互作用的模型时,它即是“空间的”。通常,这涉及一个观念,比起较远的一些位置,邻近的区域相关性更强。例如,在不动产分析中,一栋房屋的出售价格通常被认为部分地由邻里效应决定。然而,后者不可能被量化,从而只能被并入回归的误差项中。因为邻近的房屋会受到同样的邻里效应的影响,回归误差项就将是空间相关的。

空间相关的模式的形式设定需要额外的结构。有两种一般的方式可以满足要求。其一被称作地理统计模型(geostatistical modeling),即成对观测值间的相关被表示成它们之间距离的一个递减函数。这一模型形式上被并入地理统计学的半方差图(semivariogram)。另一个被称作格子或区域模型,即相关遵循一个空间的随机【Stochastic】过程,其中在某个随机变量一个给定位置的值通过一个空间加权矩阵,被表示成在其他位置的其他随机变量的一个函数。该空间加权矩阵是一个正的矩阵,其中相邻元素的相关具有一个非零的值(根据惯例,对角线上的元素被设定为0)。相邻元素的界定是相当一般的,无须根据地理上的考虑(如社会网络)。

除了误差项相关以外,空间交互作用也能被合并到一个空间滞后的因变量(或空间滞后,spatial lag)中,其本质上是因变量在邻近位置上的一个加权平均值。所得到的混合的回归—空间自回归模型(mixed regressive-spatial autoregressive model)是用于横截面背景下空间外部性和空间交互作用的最常见的模型设定。例如,在研究地方公共财政的模型中,一个地区的税率集可以

表示为邻近地区的税率的函数(一个空间滞后的因变量)。

忽视空间效应会影响最小二乘【Ordinary Least Squares, OLS】估计量【Estimator】的性质。忽视一个空间滞后会使得 OLS 的估计结果不一致,而忽视一个空间上相关的误差项会导致无效(但仍是无偏的)的估计。根据最大似然估计法则、工具变量和广义矩量法(general method of moments),已经研发了专门的估计方法。对空间自回归的诊断则是根据 Moran 的自相关统计量 I(Moran's I autocorrelation statistic),以及拉格朗日乘数原理(Lagrange multiplier principle)的应用。最新的扩展包括:以层级设定的形式,将潜变量的和计数数据的空间相关并入模型。对这类模型的估计通常需要应用模拟【Simulation】估计量,如 Gibbs 抽样【Gibbs Sampling】模拟器。

——Luc Anselin

(马妍译　赵锋校)

参考文献

Anselin, L. (1988). *Spatial econometrics*. Boston: Kluwer Academic.

Anselin, L., & Bera, A. (1998). Spatial dependence in linear regression models, with an introduction to spatial econometrics. In A. Ullah & D. Giles (Eds.), *Handbook of applied economic statistics* (pp. 237-289). New York: Marcel Dekker.

Cressie, N. (1993). *Statistics for spatial data*. New York: Wiley.

斯皮尔曼相关系数(Spearman Correlation Coefficient)

作为皮尔逊相关系数【Pearson's Correlation Coefficient】的一个特例,斯皮尔曼相关系数(r_s)是一个使用最广泛的非参数相关系数,其作用在于衡量两个定序层次变量的关系。假设有某个经理评估 n 个下属的创造力(记作 X)和生产力(记作 Y)。经理根据每个雇员的创造力和生产力,分别为每一个人评定一个等级(即 $1, 2, \cdots, n$)。假定两个变量都没有并列的等级,那么可以使用斯皮尔曼相关系数来估计创造力和生产力间的关系。

$$r_s = 1 - \frac{6\sum d^2}{n^3 - n}$$

其中 d 是 X 和 Y 的差值。这个公式实际上是从皮尔逊相关系数公式简化而来(Chen & Popovich, 2002)。

仔细考察简化的公式可以发现,所有 n 个主体上两个变量之间的差越大,即 $\sum d^2$ 的值越大,相关系数的值就越小。表 1 给出了一个关于 20 位钢琴家比赛结果的虚构案例。每位钢琴家都根据技术难度和艺术表现得到评估。假设一位裁判就每一个参赛者的技术难度和艺术表现两个方面分别确定一个 1 到 20 的等级,等级 1 表示最佳表现。假定不存在并列的等级,那么由斯皮尔曼相关系数或皮尔逊相关系数所计算出的两类成绩间的关系为 0.85。需要强调的是,如果存在着并列的等级,那么使用上文提及的简化公式就不能正确地计算斯皮尔曼相关系数。计算斯皮尔曼相关系数最简便的方法是,无论是否存在并列的等级,都用皮尔逊相关系数予以计算。

表 1 由一位裁判给出的 20 位钢琴家在技术难度和艺术表现上的等次

技术难度 (X)	艺术表现 (Y)	X^2	Y^2	XY	$d^2=(X-Y)^2$
1	2	1	4	2	1
6	4	36	16	24	4
3	6	9	36	18	9
4	3	16	9	12	1
5	1	25	1	5	16
2	5	4	25	10	9
7	7	49	49	49	0
8	9	64	81	72	1
9	12	81	144	108	9
17	10	289	100	170	49
11	8	121	64	88	9
12	11	144	121	132	1
13	13	169	169	169	0
14	15	196	225	210	1
15	14	225	196	210	1
16	20	256	400	320	16
10	18	100	324	180	64
18	16	324	256	288	4
19	19	361	361	361	0
20	17	400	289	340	9
$\sum X=210$	$\sum Y=210$	$\sum X^2=2\ 870$	$\sum Y^2=2\ 870$	$\sum XY=2\ 768$	$\sum d^2=204$

$$r=\frac{2\ 768-\frac{210\times210}{20}}{\sqrt{2\ 870-\frac{210^2}{20}}\sqrt{2\ 870-\frac{210^2}{20}}}=0.85 \qquad r_s=1-\frac{6\sum d^2}{n^3-n}=1-\frac{6\times204}{20^3-20}=0.85$$

注：皮尔逊相关系数(r)计算和展示在左边。注意，r 和 r_s 的计算结果相等。

如果将斯皮尔曼相关系数运用到定距或定比尺度测量的变量，那么变量的等级需要首先进行人工分配。表 2 列示了 10 个机构年度的专利申请数(X)和季度销量(Y)。两个变量都以定比尺度测量。在使用斯皮尔曼相关系数之前，两个变量的值都应当转换成相应的等级，其中等级 10 表示最高级别。转换后，可以运用之前描述的步骤得到斯皮尔曼相关系数。注意，在季度销量变量上有并列的等级(机构 A 和机构 F)。因此，

$$r_s=1-\frac{6\sum d^2}{n^3-n}$$

这个公式不再适用，需要通过皮尔逊相

关系数的公式来获得斯皮尔曼相关系数。以原始数值计算的皮尔逊相关系数的数值倾向于大于以转换后的等级计算的斯皮尔曼相关系数,尽管表2的示例显示了一个例外。

表2 年度专利申请数和季度销量(百万)的关系

机构	专利申请数(X)	季度销售(Y)	X 的等级(RX)	Y 的等级(RY)
A	20	5.60	5.0	6.5
B	9	1.10	2.0	1.0
C	50	6.60	9.0	9.0
D	33	7.20	8.0	10.0
E	17	4.30	3.0	4.0
F	25	5.60	6.0	6.5
G	79	6.30	10.0	8.0
H	5	2.90	1.0	2.0
I	27	4.90	7.0	5.0
J	19	3.70	4.0	3.0
相关	根据未转换的数据(X 和 Y)计算的皮尔逊相关系数是0.68		根据转换的数据(RX 和 RY)计算的斯皮尔曼相关系数是0.89	

注:机构A和机构F的季度销量具有并列的等级,等级6和等级7的均值是6.5,因此两个机构的等级均为6.5。

$\rho_s=0$ 的虚无假设检验可以计算如下:

$$t = r_s\sqrt{\frac{n-2}{1-r_s^2}}$$

当样本量是 $19 \leqslant n<30$ 时 $df=n-2$。然而,当 $n>35$ 时,$\rho_s=0$ 的检验可以通过 $z=r_s\sqrt{n-1}$ 进行更精确的计算(Kendall & Gibbons, 1990)。如果对于两个公式 n 都较小,由于带有小样本的 r_s 的抽样分布既不服从 t 分布,也不服从 z 分布,就需要使用精确的临界值。当 n 较大时,两种检验并无明显区别(Siegel & Castellan, 1988)。

——Peter Y. Chen

Autumn D. Krauss

(马妍译 赵锋校)

参考文献

Chen, P. Y., & Popovich, P. M. (2002). *Correlation: Parametric and nonparametric measures*. Thousand Oaks, CA: Sage.

Kendall, M., & Gibbons, J. D. (1990). *Rank correlation methods* (5th ed.). New York: Oxford University Press.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.

斯皮尔曼-布朗公式(Spearman-Brown Formula)

该公式由查尔斯·斯皮尔曼和威廉·布朗(Spearman & Brown, 1910)创建,用来估计多题项测量或指数【Index】的信度【Reliability】,为测量的统计理论提供了重要的基础。如果同一属性的更多指标【Indicator】用于测量时,斯皮尔曼-布朗公式也可以计算信度上期望的增量。由于具有上述两种功能,斯皮尔曼-布朗公式经常用于社会科学研究中,特别是在那些使用多题项指标的社会学、心理学和教育学有广泛的应用。

为了清晰地说明,让我们依着布朗(Brown, 1910)的研究,设定 x_1 和 x_2 是对某个具体特征 x 的两项测量,它们共同组成了一项"测验",同时,在该测验中 x_1 和 x_2 具有相等的权重。而 x_1' 和 x_2' 是同一特征的另外两项测量,两者结合在一起,给出了对特征的一个平行 x 的测验。假定所有的测量

都同样可信,或者说:

$$\sigma_{x_1} = \sigma_{x_2} = \sigma_{x_1'} = \sigma_{x_2'} = \sigma_x$$

其中,σ_{x_1}是 x_1 的标准差,并且它们的相互关系是相当的:

$$\sigma_{x_1x_1'} = \sigma_{x_1x_2'} = \sigma_{x_2x_1'} = \sigma_{x_2x_2'} = n\sigma_x^2 r$$

其中,$\sigma_{x_1x_1'}$是 x_1 和 x_1'的协方差,r 是每个测验的信度的一个量度,n 是测验的数量(在这个例子中,$n = 2$)。在实践中,如农纳利和伯恩斯坦(Nunnally & Bernstein,1994)所建议的,r 可能是构成一项测验中所有被观测的测量的平均相关【Correlation】(在这个例子中,即 x_1 和 x_2 的相关系数)。于是,可以使用一个调整的相关系数来估计指标的信度,根据布朗(Brown,1910)得:

$$\begin{aligned}
r_{SB} &= \frac{\sigma_{(x_1+x_2)(x_1'+x_2')}}{n\sigma_{(x_1+x_2)}\sigma_{(x_1'+x_2')}} \\
&= \frac{\sigma_{x_1x_1'} + \sigma_{x_1x_2'} + \sigma_{x_2x_1'} + \sigma_{x_2x_2'}}{n(\sigma_{x_1}^2 + \sigma_{x_2}^2 + 2r\sigma_{x_1}\sigma_{x_2})} \\
&= \frac{4n\sigma_x^2 r}{n(2\sigma_x^2 + 2r\sigma_x^2)} \\
&= \frac{4n\sigma_x^2 r}{2n\sigma_x^2(1 + r)} \\
&= \frac{2r}{1 + r}
\end{aligned}$$

更一般地讲,斯皮尔曼-布朗系数可以计算如下:

$$r_{SB} = \frac{nr}{1 + (n - 1)r}$$

上式中,r 是信度的一个量度(或是诸平行测验之间的平均相关),而 n 是平行测验的数目。

以下面的示例为示范:威廉·希韦尔(Sewell,1942)的兴趣在于发展出一项对农业家庭的社会经济地位的测量。从一个有123 个项目的初始清单中,希韦尔构造出一个包含了 36 个指标的指数,而他认为该指数可以测量一个家庭的地位。接着,希韦尔依据诸指标项的序号(偶数和奇数),将整个指标集分成两个"半数测验(half-tests)",并估计两个测验的相关为 0.83。使用斯皮尔曼-布朗公式,希韦尔调整了两个半数测验间的相关估计,从而产生了对他的测量信度的一个估计,而该估计将整个测验的长度纳入考量:

$$\begin{aligned}
r_{SB} &= \frac{nr}{1 + (n - 1)r} \\
&= \frac{2\times0.83}{1+(2-1)\times0.83} \\
&= 0.91
\end{aligned}$$

因为斯皮尔曼-布朗系数的范围是 0 到 1,其中 1 表示信度可能的最高水平,所以希韦尔对所得结果表示满意。然而,如果希韦尔想要增加他的测量的信度,那么斯皮尔曼-布朗公式预测,通过指标数目的倍增,使得每个半数测验成为四个平行测验中的一个(在原来假定不变的情况下),可以获得信度的增加:

$$\begin{aligned}
r_{SB} &= \frac{nr}{1 + (n - 1)r} \\
&= \frac{4\times0.83}{1+(4-1)\times0.83} \\
&= 0.95
\end{aligned}$$

相似地,斯皮尔曼-布朗"预言"公式可以用于估计一项测量到达某个要求的信度水平所需要的指标数量。

分析者应该注意,在斯皮尔曼-布朗公式的推导过程中做了两个假定:首先,假定在所有的测量中,每一测量的期望值和方差均相等。第二,正如前文所述,假定所有的

成对的测量之间的协方差均相等。换言之，所有的测量平均起来要能生成相同的均值分，应当是同等精确，还应当同等地彼此相关。这些假定——描述了“经典平行论”或一个“可交换的协方差结构”的假定（Li, Rosenthal, & Rubin, 1996）——是重要的，因为它们容许分析者用一个指数的诸构成部分之间的相关系数的一个调整的量度，来估计一个指数的信度。

然而，在社会科学中，这些条件似乎极少满足。认识到这一点，一些研究者提出一般化的斯皮尔曼-布朗公式，而新的公式可以容易地整合具有不同精度水平的测量，特别是克龙巴赫、弗雷德里克·库德和马里奥·理查森追随斯皮尔曼和布朗，已经逐步建立了测量的统计理论的重要基础。参见词条 Cronbach α【Cronbach's Alpha】。

——Karen J. Long

（马妍译　赵锋校）

参考文献

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, *12*, 296-322.

Li, H., Rosenthal, R., & Rubin, D. (1996). Reliability of measurement in psychology: From Spearman-Brown to maximal reliability. *Psychological Methods*, *1*(1), 98-107.

Nunnally, J., & Bernstein, I. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Sewell, W. H. (1942). The development of a sociometric scale. *Sociometry*, *5*(3), 279-297.

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, *12*, 271-295.

设定（Specification）

模型设定是指，对自变量【Independent Variable】生成因变量【Dependent Variable】的过程的描述。因此，设定包括：自变量（及因变量）的选择，以及连接自变量和因变量的函数形式的选择。设定也可能包括，同模型【Model】的随机【Stochastic】构成相关的所有假设【Assumption】。由此，设定先于估计【Estimation】。人们熟知的一个模型设定的简单例子是简单线性模型，其中假定某个变量 Y 是一个随机【Random】变量 X 的线性函数【Linear Function】，且有一个随机扰动项 ε：

$$Y = \alpha + \beta X + \varepsilon \qquad (1)$$

还有许多替代的设定，可以关联 Y 与 X，如：

$$Y = \alpha + \beta X^2 + \varepsilon \qquad (2)$$

或

$$Y = \alpha + \beta X^3 + \varepsilon \qquad (3)$$

一旦设定了一个模型，一项估计技术就只是一个确定模型中参数值——根据模型设定，确定最能拟合观测数据的——工具。然而，没有正确的模型设定，任何估计实践都只是不能做出有效统计推论【Statistical Inference】的曲线拟合。为了从观测的样本【Sample】数据推断总体【Population】，就需要我们假设，总体中所有的 Y 值都经由产生样本的观测值的同一过程产生。换言之，我们必须假定，设定是正确的。在著名的高斯-马尔科夫定理【Gauss-Markov Theorem】中，这一般被表述为“模型是正确设定的”。

模型设定应当根据某些相关理论。如果模型的被估参数具有统计显著性【Statistical Significance】，可能会认为模型是对理论的确证。但事实上，它仅是对设定的

模型相对于零模型(一个假定模型的自变量和因变量没有关系的模型)的一个技术上的检验。如果我们对上面的三个模型中的任意一个进行估计,并且能够在95%的水平上拒绝$\beta=0$的虚无假设【Null Hypothesis】,我们仍然不能说,我们所估计的设定就是正确的,同时另外两个提议的设定都是错误的。

尽管有无限多种形式的设定错误(specification error),但只有几种常见的设定错误需要留意。其中最普遍的是遗漏变量【Omitted Variable】问题。如果设定的模型遗漏了一个自变量,而这个变量实际上是生成数据的真实模型的组成部分,那么该模型即是误设的。这类设定错误的后果很严重。即使在最小二乘【Ordinary Least Squares】的简单情形和线性模型中,其他诸变量的估计参数可能是偏倚的,进而我们可能作出无效的统计推论。

没有任何检验可以确定我们有绝对正确的模型设定。然而,我们通常可以用统计检验,来比较我们的设定和诸备择设定。最简单的检验涉及那些嵌套在其他模型中的模型设定。如果一个设定与另一设定有同一函数形式,且前者的变量是后者变量的子集,那么该设定即嵌套在另一模型设定中。例如,模型$Y=\alpha_0+\beta_1X_1+\varepsilon$嵌套在模型$Y=\alpha_0+\beta_1X_1+\beta_2X_2+\varepsilon$中,但是模型$Y=\alpha_0+\beta_3X_3+\varepsilon$并未嵌套在模型$Y=\alpha_0+\beta_1X_1+\beta_2X_2+\varepsilon$中。就线性模型而言,通常用线性约束的$F$检验【$F$ Tests】(或邹氏检验【Chow Tests】),检验嵌套模型中一个设定对另一个设定的有效性,而对数似然比检验则适用于最大似然【Maximum Likelihood】模型。在所有计量经济学教材中,此类检验都有叙述。在两种情况下,人们先估计两个模型中的每一个,再根据每个模型的拟合结果,生成一个检验统计量。然后,可以运用经典的假设检验,而这让人们可以拒绝一个模型而接受另一个。在用F检验比较两个线性设定(其中一个嵌套在另一个中)的情况下,检验统计量如下:

$$F=\frac{(R_{UR}^2-R_R^2)/m}{(1-R_{UR}^2)/(n-k)} \qquad (4)$$

该统计量服从自由度【Degrees of Freedom】为$(m,n-k)$的F分布,其中m是线性约束的数量,n是观测值的数量,而k是等式右边变量的总数。

非嵌套设定的检验通常更难进行。然而也存在相应的检验。Cox检验可以进行两个非嵌套模型的统计检验;而Vuong检验也可以在非嵌套模型之间进行比较。

在竞争性模型设定中进行选择的另一种方式是使用模型选择准则。模型选择准则(criteria)并不提供一项统计检验,来表明偏好某个模型胜于另一个。它们是从诸竞争模型中选出最适当模型的公认标准。现实中,模型选择准则通常是拟合优度【Goodness-of-Fit】(在所有条件相同的前提下,我们偏好能解释更多方差的模型而非解释更少的模型)和简约(几乎所有的选择准则都科罚那些包括了额外变量的模型设定)之间的一种权衡。常用的准则包括Akaike信息量准则(Akaike Information Criterion, AIC)、Schwartz准则(Schwartz Criteria, SC),以及Bayes信息量准则(Bayes Information Criteria, BIC)。就线性模型而言,AIC的计算很简单:

$$\mathrm{AIC}(k)=e^{\frac{2k}{n}}\frac{\sum e_i^2}{n} \qquad (5)$$

通常会给出它的对数形式:

$$\ln \mathrm{AIC}(k)=\frac{2k}{n}+\ln\left(\frac{\sum e_i^2}{n}\right) \qquad (6)$$

那些准则虽然是合理的,但终归还要涉及随意的权衡。在缺乏真正统计检验的前提下,研究者不能就他或她有多少信心在诸

模型中做出了正确的选择给予任何类型的概率陈述。我们拥有正确模型的假设是计量经济学【Econometrics】中最有争议的争论之一。它催生了利默(Leamer,1983)的著名论文“让我们把骗子赶出计量经济学(Let's Take the Con out of Econometrics)”,在文中,他建议抛开假设,代之以检查右侧变量的所有合理的组合,并报告参数的范围。由于这类范围一般很大,所以这是一个没有实践者采纳的建议。但是,这里要建议研究者检验模型设定,以达到使扰动保持稳健的目的。

——Jonathan Nagler

(马妍译　赵锋校)

参考文献

Leamer, E. E. (1983). Let's take the con out of econometrics. *American Economic Review*, *73*(1), 31-43.

频谱分析(Spectral Analysis)

频谱分析是一种识别时间序列【Time-Series】数据的循环成分的方法。循环是那些常可表示为一个三角正弦函数、余弦函数或者正弦曲线的波形的周期事件,其图形和数学的表示如图 1 所示。

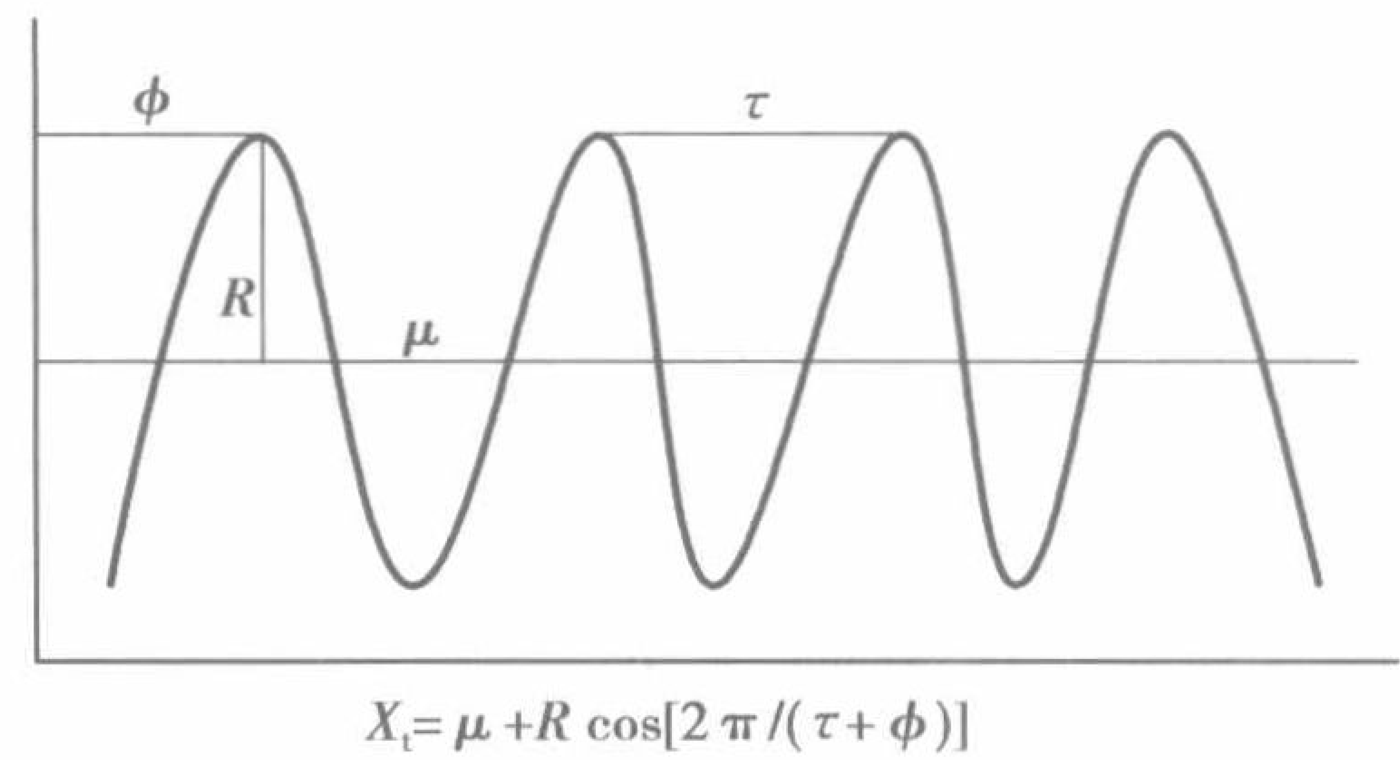


图 1　频谱分析

其中τ表示波形的周期,或者从循环的一个顶点到另一个顶点的距离;μ 表示波形最高点和最低点的均值;ϕ 表示波的初始阶段;而 R 表示振幅,或者从波形的顶点到均值线的距离。

正弦曲线是周期分析中的一个重要成分,因为它常用于为这些循环的形状建模【Modeling】。如果存在复合的波形,研究者和统计软件包可以使用几个正弦曲线成分为循环建模。即便循环实质性地背离正弦曲线,为数据建模仍是可能的,不过这类方法在数学上很复杂而且比本文提出的模型更难以处理。

一旦数据中的时间趋势被移除,即通过将一个时间序列上那些不能归之于数据中的一个线性或曲线【Curvilinearity】趋势的残余方差分解出来,就可以近似地得到时间序列数据的循环性质。总平方和【Sum of Squares】被分入组数为 $N/2$ 的平方和成分,每个成分的自由度【Degrees of Freedom】为 2。为了创建强度估计值,从计算中得出的估计值被转成由每一周期部分解释掉的平方和。

为了收缩常常伴随着强度估计值的较宽的置信区间【Confidence Intervals】,频谱分析需要平滑强度估计值,从而创建一个能谱。平滑【Smoothing】程序是用几个邻近的频次计算强度估计值的加权平均。由于不同平滑技术的使用,在包含于加权平均中的邻近频次数量(或宽度)上的变化,以及所应用的权数的变化,都能产生可靠的数据的循环性质的估计值。例如,丹尼尔(相等的权数)平滑技术,即用宽度(或 M)3 来代替频次 i 的强度估计值,亦即使用三个坐标 $i-1$, i 和 $i+1$ 的均值代替频次 i 的强度估计值。相等的权数(或 $1/M$)被用于新的估计值,以获得循环的一个平滑估计值,即

(1/3)×在频率($i-1$)处的强度+(1/3)×在频率 i 处的强度+(1/3)×在频率($i+1$)处的强度。值得注意的是,随着邻近频次数量的增加,置信区间的宽度变窄。过度平滑可能扭曲数据,并且难以对数据中周期的长度(或τ)作出明确的判断。

一旦计算出了能谱,就可以用它除以能的总量或方差的总量,从而产生一个频谱密度函数。通常,结果会被绘制在图上,其中 y 轴代表频谱密度,x 轴代表频率($1/\tau$)。一个以 4 年为周期的数据会在一个频率的 1/4 或 0.25 的位置上标记它的峰值,而这可以清楚地表现在一张平滑图上。根据卡方分布【Chi-Square Distribution】,通过置信区间的估计,可以确定统计显著性。

$$下限 = [edf^* s(f_1)]/X^2.99$$
$$上限 = [edf^* s(f_1)]/X^2.01$$

其中,edf 是平滑前初始自由度(即 2)和平滑技术的宽度的乘积,而 $s(f_1)$ 表示一个给定频率的频谱估计。如果置信区间与能的平均水平(或者频谱估计的均值)重叠,那么我们可能得出这样一个结论:峰值并不显著地不同于基于机会所能期望的峰值。替代方案是,针对未平滑的峰值,用所有最大的平方和(最大峰值的平方和)或所有最大的强度估计值除以总平方和之和($\sum$ 总平方和)。

——Erika Moreno

(马妍译　赵锋校)

参考文献

Granger, C. W. J. (1964). *Spectral analysis of economic time series*. Princeton, NJ: Princeton University Press.

Warner, R. M. (1998). *Spectral analysis of time-series data*. New York: Guilford.

球形假定(Sphericity Assumption)

球形假定是针对受试者内设计【Within-Subject Design】使用单变量(或者混合模型)方差分析【Analysis of Variance】进行统计推论所必需的。简略地说,球形假定假设重复测量的任意线性组合的方差【Variance】是相等的。特别地,各处理水平间的差值具有相等的方差。如果这一假定未被满足,就有必要采用统计修正,或者使用多元方差和协方差的分析【Multivariate Analysis of Variance and Covariance】。

用一个例子来阐明该假定。设想有 3 个人接受了 3 种治疗,得到了表 1 所示的数据(和差值)。

表 1

治疗 1	治疗 2	治疗 3	1—2	1—3	2—3
5	7	2	−2	3	5
3	7	4	−4	1	3
6	8	6	−2	0	2
方差			1.33	2.33	2.33

如要合理地将数据当作一个两因素设计(即双因素方差分析【Two-Way ANOVA】),处理间差值的方差就必须是相等的。就上述假设的例子而言,正如最后 3 列所示,球形假定未予满足,因为差值的方

差不相等。注意,如果只有两种处理,根据定义,球形假定得到满足。

错误地忽视对球形假定的背离会导致一个检验效力【Power of a Test】的损失。当球形假定未被满足时,犯第一类错误【Type Ⅰ Error】的实际概率比用标准计算暗含的概率高。换句话说,研究者会冒风险作出不为数据支持的统计推论,因为相比于正常的可能性,不存在处理差异的虚无假设会更经常地被拒绝(但是更可能的量的大小是未知的)。注意,如果无差异的虚无假设没有被拒绝,就没必要作出修正,因为同这一概率相比,未调整的检验统计量是有偏的。

如果球形假定未被满足,研究者有两种选择:一种选择是人们可以试着调整检验统计量,从而消解假定的背离。有许多调整方法可供使用,这些方法在其他地方有具体阐述(如 Maxwell & Delaney, 1990)。另一种选择是采用方差和协方差的多元分析。除非样本量非常小,多数人都推荐使用多元方法。

——Joshua D. Clinton

(马妍译　赵锋校)

参考文献

Bogartz, R. S. (1994). *An introduction to the analysis of variance*. Westport, CT: Praeger.

Field, A. (1998). A bluffer's guide to ... sphericity. *British Psychological Society: Mathematical, Statistical & Computing Newsletter*, *6*, 13-22.

Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparison perspective*. Belmont, CA: Wadsworth.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). New York: McGraw-Hill.

样条回归(Spline Regression)

样条回归是在称作分段节点的具体的连接点上连接在一起的一系列回归【Regression】线段。根据定义,在一条样条组成的线上,不存在突然的中断。然而,在它们的节点处,样条可能突然改变方向(即斜率【Slope】),或者,像曲线一样,在斜率的变化率上发生更加细微的变化。

样条线被用来取代二次方【Quadratic】、三次方或者更高阶的多项式【Polynomial】回归曲线,因为样条线更灵活,且较少受多重共线性【Multi-collinearity】的影响,而这种影响正是多项式回归的顽疾。样条回归可生成比中断【Interrupted】回归更平滑的曲线,而后者则容许在线上发生中断或跳跃。

若政策上的一项变化,在政策引入的时刻,仅影响了回归线的斜率,而没有导致回归线的一个中断或跳跃,这时就可以使用样条。例如,对更多的住宅隔热项目,宣布一项能源课税抵扣政策会减低能源消费曲线的斜率,而不导致能源消费上的一个突然减少。税收安排通常分为不同的纳税等级,当随着收入的增加,代表纳税额的回归线从一个纳税等级过渡到另一个纳税等级时,这条回归线不会发生突然的变化,而是在这些样条的节点处发生斜率的变化。

诸段的线性样条好像带有钮结(kinks)的直线。这些钮结即节点位置斜率的变化。当年龄小于等于 21 岁时,定义 $D=0$,由此,样条 D(年龄-21)= 0 表示年龄为 0 到 21 岁的回归线;当年龄大于 21 岁时,定义 $D=1$,

由此，样条 D(年龄-21) = 1, 2, 3,…表示年龄为22,23,24等的回归线；因此，当年龄超过21岁时，直线在钮结处(即在样条的节点位置)转向一个新的斜率，而不发生回归线的跳跃或中断。

如果事先知道样条节点数和它们的位置，样条回归是非常容易估计的。宣布的政策变动和纳税等级是最好的例子。用最小二乘【Ordinary Least Squares】可以容易地估计此类样条回归。

然而，有时节点位置无法准确获知。例如，人们的收入通常随着年龄的增加而上升，然后，在某个未知的点开始普遍地下降。可以更精确地估计出这个点吗？这类未知的样条节点的位置可以用非线性【Nonlinear】最小二乘予以估计。

如果样条节点的数量未知会如何？可以使用逐步回归【Stepwise Regression】来搜寻未知样条节点的数量和位置。一份潜在的样条节点清单，作为纳入回归方程的可能变量，以变量的形式被创建出来。如果在所有设定的样条节点的备选集中，一个样条的节点位置和校准度(斜率、斜率的变化率等)给出了最大限度的统计显著【Statistical Significance】，这一特定的样条节点即被引入回归方程。

分段变量可能是时间变量或者某些截面【Cross-Sectional】变量，如个人的年龄或生产水平。样条可用于估计三维平面。例如，在控制了多元列表卡上的诸变量(卧室、浴室、学区等)后，一个三维的样条回归能用于估计一栋住宅的位置价值。

——Lawrence C. Marsh

(马妍译　赵锋校)

参考文献

Marsh, L. C., & Cormier, D. R. (2001). *Spline regression models*. Thousand Oaks, CA: Sage.

Pindyck, R. S., & Rubinfeld, D. L. (1998). *Econometric models and economic forecasts* (4th ed.). New York: McGraw-Hill.

折半信度(Split-Half Reliability)

在经典的测验分理论中，有 $X=T+e$(其中，X=观测得分，T=真实分数，e=误差)。其中有些误差是减低测验效度的常量，而有些误差是导致得分不一致的变量。信度是测验免于变量误差影响的程度。形式上，给定 $\sigma_x^2=\sigma_t^2+\sigma_e^2$，信度是真实分数的方差对误差分数的方差的比，即

$$\frac{\sigma_t^2}{\sigma_x^2}=\frac{\sigma_x^2-\sigma_e^2}{\sigma_x^2}\text{ 或 }=1-\frac{\sigma_e^2}{\sigma_x^2}$$

因为测验题项必定是被测领域的一个样本，所以它们本身构成了变量误差的一个来源。可以通过比较来自不同时点上两个测验的等价(备择)形式的得分，估计测验题项对变量误差的影响，但是，第二次测验的成绩可能受第一次测验的影响。折半信度只要求一次测量实施，因此参与者不会知道比较的发生。将所有题项分成两个等价的部分，再计算两部分得分的相关系数(皮尔逊积矩系数，r)，即可得到折半信度。有多种方法可以用来分拆题项，但是其中有些分拆方法无法形成等价的题项集合。例如，题目难度可能在两个随机选择的题项集合中不同，也可能在两个由测验的第一半部分和第二半部分造成的题项集合中不同。两半部分的得分也可能由于实践或疲劳而不

同,由此人为地降低了折半估计。

由于任一种折半信度的值都不是唯一的,所以至少应当计算两种类型的折半信度。甚至,存在着用以获得最大的真实信度的算法。然而,最常见的技术是将奇数项和偶数项进行分组。这种技术可以克服疲劳和实践效应,同时,如果题项难度随测验而增加,还可以使得两组题项的难度取得对等。遗憾的是,如果题项的不同种类交替出现,如那些要求回答真/假,或是/否的问题,奇/偶分拆法未必能保证题项的等价。最佳的方法是通过题目的内容、难度和回答类型对题目进行匹配,系统地创建两个等价的题项集合(即平行分拆)。两个题项集合的均值和方差应相等。

因为测验信度是一个题项数量的正的递减函数,折半 r 将低估整体测验信度。因此,用斯皮尔曼-布朗【Spearman-Brown】校正公式 $r'=2r/(1+r)$ 来对它进行修正。整体测量信度也可以用如 Rulon 公式 $r''=1-\sigma_d^2/\sigma_x^2$(其中 σ_d^2 是测验的两部分的分数差值的方差)直接进行估计。显然,这反映了信度的经典定义($1-\sigma_e^2/\sigma_x^2$, 其中 $\sigma_e^2=\sigma_d^2$)。当 $\sigma_1^2=\sigma_2^2$ 时,它将产生与斯皮尔曼-布朗校正同样的值;但是当 $\sigma_1^2\neq\sigma_2^2$ 时,它的值低于斯皮尔曼-布朗的校正值。

折半信度不适用于快速完成的测验,因为人们在两个题项集合上的得分或者都很低,或者都很高,从而导致一种虚假的高度相关。它也不适用于多维测验,因为不同的部分可能不相关。对于一个同质测验而言,它则是一个方便的技术。尽管不存在判断可信的通行标准,但是折半信度(源于测验题项的误差)通常应超过 0.80 或者 0.85,而且它通常也比再测信度【Test-Retest】(源于时间的误差)和复本信度【Alternate Form】(源于不同的题项和时间的误差)高。

例 1:一个由 148 人参加的,有 24 道题的测验(均值 $M=11.20$,方差 $s^2=20.20$)被分为两个题项集合,每个集合中有 12 道题。两个题项集合的均值(5.38 和 5.89)和方差(7.12 和 5.57)没有显著差异,表明两个题项集合是等价的。两集合的 r 值为 0.626, $s_d^2=4.81$。使用斯皮尔曼-布朗校正公式 $r'=2\times0.626/(1+0.626)=0.770$;使用 Rulon 公式 $r''=1-4.81/20.20=0.762$。两个值都表明折半信度低于要求的信度值。

例 2:一个由 59 人参加的,有 20 道题(5 分制量表)的测验(均值 $M=56.61$,方差 $s^2=240.449$)被分为奇数题项集和偶数题项集。两个集合的均值(28.37 和 28.24)和方差(68.96 和 59.25)没有显著差异,表明两个集合是等价的。两组的 r 值为 0.878,$s_d^2=15.98$。使用斯皮尔曼-布朗校正公式 $r'=2\times0.878/(1+0.878)=0.935$,使用 Rulon 公式 $r''=1-15.98/240.449=0.934$。两个值都表明折半信度非常理想。

——Stuart J. McKelvie

(马妍译　赵锋校)

参考文献

Callendar, J. C., & Osburn, H. G. (1977). A method for maximizing split-half reliability coefficients. *Educational and Psychological Measurement*, *37*, 819-825.

Cronbach, L. J. (1943). On estimates of test reliability. *Journal of Educational Psychology*, *34*, 485-494.

Cronbach, L. J. (1946). A case study of the split-half reliability coefficient. *Journal of Educational Psychology*, *37*, 473-480.

Gulliksen, H. (1950). *The theory of mental tests*. New York: Wiley.

Lord, F. M. (1956). Sampling error due to choice of split in split-half reliability coefficients. *Journal of Experimental Education*, *24*, 245-249.

裂区设计(Split-Plot Design)

裂区设计是一种研究设计,其中实验【Experiment】的单元(如消费者、父母、学生)接受或被分配到一个因子(如因子B)的所有处理中,而只接受另一个因子(如因子A)的一种处理。在裂区设计中,因子A被称为"受试者间"因子,而因子B被称为"受试者内"因子。例如,一个人想要研究互联网搜索策略(因子A)对于大学生在健康和历史两个领域内信息搜索效率(自变量)的影响(因子B的水平)。在一个州立学院随机抽取了100名大学新生,并将他们随机分入因子A的两个条件中:50名学生分到"线性搜索"的条件下,剩下50人分入"非线性搜索"的条件下。所有受试者都被指示就两个主题(健康和历史)搜索适当的信息。搜索的成功度由三个标准来判断:完整性、相关性和速度。下面的表格例示了这一研究设计:

		因子B:主题			
		受试者	健康	历史	原始均值
因子A:互联网搜索策略	处理1 线性搜索	S_1 ⋮ S_{50}	$\bar{Y}_{11}$	$\bar{Y}_{12}$	$\bar{Y}_{1.}$
	处理2 非线性搜索	S_{51} ⋮ S_{100}	$\bar{Y}_{21}$	$\bar{Y}_{22}$	$\bar{Y}_{2.}$
列均值			$\bar{Y}_{.1}$	$\bar{Y}_{.2}$	总均值 $=\bar{Y}_{..}$

在表格和研究设计中,因子A和因子B都是完全交叉的。由于两个因子的交叉性质,这个设计是一个技术上的裂区因子(或SPF)设计。而且,分配到因子A的每种条件的受试者的数目是相等的,这使得设计成为一个平衡的SPF设计。否则,就是一个不平衡的SPF设计。因为因子A有两个层次(线性搜索策略与非线性搜索策略),因子B也有两个层次(健康与历史),这个设计在教科书中被称为$SPF_{2.2}$设计。受试者间因子的层次数目总是在点号(.)之前写出来,受试者内因子的层次数目在点号之后写出来都是下标。裂区设计的变化样式有更复杂的设计形式,例如$SPF_{pq.r}$设计,有两个受试者间因子和一个受试者内因子;$SPF_{p.qr}$设计有一个受试者间因子和两个受试者内因子;$SPF_{pq.rt}$设计有两个受试者间因子和两个受试者内因子等。就这些变化样式的细节,读者应查阅基尔克(Kirk,1995),以及马克斯韦尔和德莱尼(Maxwell & Delaney,1990)的著作。

请注意,在上面的SPF设计中,受试者(在表格中表示为S)在整个研究中是嵌入在因子A(或受试者间的因子)的处理条件中的。这样,一个SPF设计也是一个嵌套设计,因为受试者被认为是嵌套在因子A(或受试者间因子)的层次中的。而且,一个SPF设计也是一个区块设计,因为主体在因子B(或受试者内因子)的所有层次中受他们自己的控制;于是,根据随机区块设计的原理,他们也被认为是"区块"。

裂区设计是社会科学中最流行的研究设计之一。它起源于农业研究,一块块的土地被称为区。区被分到次区中以调查化肥、土壤条件和这类处理的影响。林德奎斯特(Lindquist,1953)在心理学和教育学领域促进了这种设计的普遍使用,并重新命名它为混合设计【Mixed Design】。混合设计这个词反映出设计中受试者间因子和受试者内因子的独特混合。裂区设计也被作为模型

Ⅲ【Model Ⅲ】设计(Hays, 1994),在这个意义上,区块(或上面的例子中的受试者)通常被作为一个随机因子处理,从而因子 A 或因子 B(或两者)是一个混合因子。在同一个 SPF 设计中包含随机和固定因子,要求统计模型是一个混合模型,如下所示:

$$Y_{ijk} = \mu + \alpha_j + \pi_{i(j)} + \beta_k + (\alpha\beta)_{jk} + \beta\pi_{ki(j)} + \varepsilon_{ijk}$$
$$(i = 1,\cdots,n;j = 1,\cdots,p;k = 1,\cdots,q)$$

其中:

Y_{ijk}是第 i 个区块(也是观察和分析的单元)在因子 A 的第j个水平上和因子 B 的第 k 个水平上的得分。

μ 是所有观测总体的总均值,亦即一个常数。

α_j 是因子 A 的第j个处理条件的效应;代数上,它等于总体均值($\mu_{j.}$)和总均值(μ)的离差。对所有在第j个条件下的观测值的因变量得分而言,它是一个常数,其约束条件是所有因子 A 的处理条件 α_j的和等于零。

$\pi_{i(j)}$是嵌套在因子 A 的第j个条件下,与每个区块 i(例如上面图表中的主体)相关的总体效应。它的潜在总体分布被假定为是正态的,并在每个区块中有相等的方差。

β_k 是因子 B 的第 k 个处理条件下的效应。代数上,它等于总体均值 ($\mu_{.k}$)同总均值(μ)的离差。对于第 k 个条件下的所有受试者的因变量得分而言,它是一个常数,且服从约束条件,对于因子 B 的所有处理条件,所有 β_k 之和为零。

$(\alpha\beta)_{jk}$是由于因子 A 的j层次和因子 B 的 k 层次的联合效应。它服从于约束条件,在因子 A 和因子 B 的所有处理条件下,所有的$(\alpha\beta)_{jk}$的和为零。

$(\beta\pi)_{ki(j)}$是因子 B 的 k 层次和区块 i 的联合效应,两者都嵌入在因子 A 的第j个条件中。假定 $(\beta\pi)_{ki(j)}$的潜在分布是正态分布,同时在每个区块和因子 B 的每个层次上有相等的方差。此外,还假定$(\beta\pi)_{ki(j)}$统计上独立于效应 $\pi_{i(j)}$。

ε_{ijk}是与嵌套在因子 A 的第j个层次和因子 B 的第 k 个条件相关的随机误差效应。它是一个随机变量,在潜在总体中呈正态分布,且独立于效应 $\pi_{i(j)}$。

要检验因子 A 的效应、因子 B 的效应,或 A 和 B 的交互效应的统计假设,需要两个误差项。第一个误差项用于检验受试者因子(即 A),可以在忽略因子 B 的条件下,计算因子 A 的各层次间的平均的(上列中是两个层次)均平方而得到。第二个误差项用于检验受试者因子(即 B),以及它与受试者间因子(即 A 与 B)的交互效应。通过计算与 B 联合在一起的 A 的各层次的平均的均平方,可以得到第二项误差。技术上,它是因子 B 以及那些被包括在因子 A 的每个层次内的所有区块的汇聚的交互效应。所有的统计检验都是在相等方差假定下,通过 F 统计来实现的。受试者内因子的 F 检验,以及它同受试者间因子的交互效应的 F 检验需要另一个额外的假定,即受试者内得分的方差和协方差的球形假定。球形假定是使得 F 检验有效的必要且充分条件(Huynh & Feldt, 1970)。

——Chao-Ying Joanne Peng

(王玥译　赵锋校)

*也可参见重复测量设计【Repeated Measures】、嵌套设计【Nested Design】、区组设计【Block Design】。

参考文献

Hays, W. L. (1994). *Statistics*. Fort Worth, TX: Holt, Rinehart and Winston.

Huck, S. (2003). *Reading statistics and research* (4th ed.). New York: Pearson Education Inc.

Huynh, H., & Feldt, L. S. (1970). Conditions under

which mean square ratios in repeated measurements designs have exact *F*-distributions. *Journal of the American Statistical Association*, *65*, 1582-1589.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Belmont, CA: Brooks/Cole.

Lindquist, E. F. (1953). *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin.

Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparison perspective*. Belmont, CA: Wadsworth.

S-Plus

S-Plus 是一款以通用为目的,并依赖于面向对象编程的统计软件包。它是一款强大且灵活的计算机程序,易于根据个人目的进行定制。更多信息可参见相关网站。

——Tim Futing Liao

(马妍译　赵锋校)

离散(Spread)

数据的离散是指数据是如何分布的。离散的量度指标包括全距【Range】、方差【Variance】和标准差【Standard Deviation】。

——Tim Futing Liao

(马妍译　赵锋校)

* 参见**离散趋势**【Dispersion】。

SPSS

SPSS(社会科学统计软件包)是一款以通用为目的的软件包,不仅为社会科学家所广泛运用,还能满足他们诸多计算需求。更多信息可参见相关网站。

——Tim Futing Liao

(马妍译　赵锋校)

虚假关系(Spurious Relationship)

虚假关系是指,尽管两个或两个以上的变量是相关的,但这些变量的关系不是因果相关。相关【Correlation】分析仅仅确立协变关系,即两个或两个以上的现象一起发生变化的程度;而因果关系【Causality】要推论,一个现象引起了另一个现象。尽管社会科学家对相关感兴趣,但是大部分社会科学家也想推论因果关系。例如,尽管我们知道吸烟与肺癌相关,我们还想推论出吸烟引发肺癌的可能性。但是当寻求因果推论时,社会科学家必须要小心虚假关系的可能性,即那些隐含了虚假的因果关系的相关关系。

设想我们要估计 r_{xy} 的相关系数,其中 x 是个人的婚姻状态,y 是个人每年消费的红酒数量。我们发现在婚姻状态和红酒消费之间存在很高的正相关。如果我们要推论

因果关系,我们可能假定消费红酒引发人们结婚。我们也可以假定婚姻驱动个人去喝红酒。这些解释在逻辑上都没有道理。很可能是红酒消费与婚姻状态之间的关系是虚假的,而第三个变量才是引发这两个变量正相关的变量。

虚假关系意味着随着第三个变量的引入,前两个变量之间的关系会消失。理论或逻辑应该明确说明第三个变量可以解释前两个变量之间观察到的相关关系。就上面的例子而言,我们怀疑变量 z,即年龄,可能同时影响婚姻状态和红酒消费。我们发现在婚姻状态和年龄之间有很高的正相关,r_{xz};同时,红酒消费和年龄之间也有很高的正相关,r_{yz}。这些强相关关系提供了最初的证据,证明年龄变量解释了婚姻状态和红酒消费之间的关系。为了检验虚假关系,我们计算控制年龄变量时婚姻状态和红酒消费之间的偏相关【Partial Correlation】系数 $r_{xy\times z}$,从而有 $r_{xy\times z}=r_{xz}\times r_{yz}$。我们发现 $r_{xy\times z}$ 接近于零,而这意味着当控制了年龄变量后,婚姻状态与红酒消费之间的关系消失了。于是我们推断,婚姻状态与红酒消费之间的关系来自于年龄变量上变差的联合效应。换言之,婚姻状态与红酒消费之间的关系可能是虚假的。

因为 $r_{xy\times z}$ 的值接近于零,我们也可能得出结论,z 是 x 和 y 之间的中介变量【Mediating Variable】。在上面的例子中,这意味着婚姻导致人变老,而变老则引起更多的红酒消费。对于我们的例子来说,这是一个荒谬的结论,但是在其他情况下,可能没有这么清楚地区分。实践中,只能由研究者就因果关系的方向作出逻辑假定,再使用这些假定来推论 x 和 y 是虚假相关的,抑或 z 是一个中介变量。统计证据本身无法区分这两种可能性。

——Megan L. Shannon

(王玥译　赵锋校)

参考文献

Blalock, H. M., Jr. (1979). *Social statistics* (2nd ed.). New York: McGraw-Hill.

Simon, H. A. (1954). Spurious correlation: A causal interpretation. *Journal of the American Statistical Association*, *49*, 467-479.

稳定系数(Stability Coefficient)

在社会科学中,稳定系数是一个术语,用来指当被测主体和测量工具都确切地保持相同时,得自于时间 1 的测量结果同得自于时间 2 的测量结果的相关度。如果系数是高的[且没有工具反应性(instrument reactivity)出现],那么这一般地指示了测量的信度【Reliability】和响应的连续性(response continuity)。如果系数低,则很难得出一个正确的解释。

个人历史或行为的测量有低的稳定系数,这无疑来自测量的不可信性。如果回答人第一次报告了某些特征的个人背景或过往行为,然后又否定了他的报告,那么人们不能将这一变化解释为“不连续性”。然而,就态度和意见而言,低的系数缺少确定的意义。如果测试与再测的间隔短暂,那么这一系数通常会被当作不可信性的一个指标(Carmines & Zeller, 1979; Litwin, 1995);如果间隔较长,通常则将其当作不连续性的一个指标(Converse & Markus, 1979)。两个解释都不具有不证自明的正确性,因此,将一个稳定系数的可信性部分

和连续性部分分解开可能既复杂又重要(Achen, 1975)。实际上,如果不知道测量的不可信性水平,又不根据不可信水平加以修正,就不可能确定连续性的精确估计值。

一个仅有几天或几周时间间隔的再测试【Test-Retest】显然会增强如下假设,即低的稳定性是由测量问题引起的,而非态度变化的结果。然而,即便如此,其他因素也可能发生干扰。例如,在动荡的时期,有可能存在相当程度的真实意见的短期混乱。同时,无论在动荡的时期还是在和平的时期,就不熟悉或考虑欠周的话题,准确地陈述意见都会显出实质性的短期的不稳定性。态度的瞬间变化不代表真正的测量误差,尽管人们抱持着一种"态度"定义,强调其具有持久的性质,而将这类瞬间变化视作"非态度"。

在估计连续性时,使用相关【Correlation】系数,无论是积矩的(product-moment)抑或秩序的(rank-order),有时都会是有问题的。就标准化的数据和排序的数据而言,都可能存在测量时间1和测量时间2之间的系统变化,而这些变化没有反映在相关系数上。响应的分布可能朝上移动或朝下移动,也可能更加集中或更加分散,与此同时,稳定系数则始终保持较高。基于这个原因,就有必要通过检视定距数据的散点图,或排序数据的列联表,进一步审查相关关系,找出分布变化的迹象。

——Douglas Madsen

(马妍译　赵锋校)

参考文献

Achen, C. H. (1975). Mass political attitudes and the survey response. *American Political Science Review*, *69*, 1218-1231.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage.

Converse, P. E., & Markus, G. B. (1979). Plus ça change ...: The new CPS Election Study Panel. *American Political Science Review*, *73*, 32-49.

Litwin, M. S. (1995). *How to measure survey reliability and validity*. Thousand Oaks, CA: Sage.

稳定人口模型(Stable Population Model)

当前的人口学模式保持了一个较长时期后,其长期的隐含意义是什么?艾尔弗雷德·洛特卡从种群更新的生物学模型到年龄别率(age-specific rates)的领域的拓展给出了许多可能的答案(Lotka, 1939/1998)。当下述非常接近于许多历史人口总体的长期条件的三项特征存在时,一个具有恒定的出生率和年龄分布的稳定人口总体即出现了,并在较长时期内持续存在:

1.年龄别生育率是恒定的;

2.年龄别死亡率(即生命表【Life Table】)是恒定的;

3.年龄别迁移率为零。

潜在于稳定人口总体之下的简单人口学属性构成了人口预测和人口学成分(年龄分布、死亡率和生育率)分析的基础。其特别的重要性在于遍历性(ergodicity)特征,即长期的年龄分布、出生率、死亡率以及稳定人口的增长率都取决于它们当前的生育率和死亡率模式,而与当前年龄分布无关。

一个稳定等价的人口总体的长期规模、人口分布和生命率可以从当前的生命率来预测(更多内容参见人口学方法

【Demographic Methods】)。在每一年龄上的人口比率（${}_nK_x^{\text{stable}}$）可以用生命表的生存函数（${}_nL_x$）、出生人数（B）和人口增长率（r）来表示：

$$ {}_nK_x^{\text{stable}} = B{}_nL_x e^{-r \cdot x} $$

总的出生人数 B 会随初始的人口年龄分布变化，但是长期的粗出生率和每年的出生数量的增长仅是生育率和妇女死亡率模式的函数。内在增长率（intrinsic growth rate）r，或者当前生命率中隐含的稳定增长率可用于对净再生产率（net reproduction rate，NRR）做直接且简单的估计，其中 NRR 指基于当前的出生率和死亡率，每个妇女平均生育的女儿数：

$$ \text{NRR} = e^{rT} $$

其中 T 是世代之间的平均间隔时长。

图 1 通过一个简单的假设人口预测描绘了稳化过程（参见 Leslie 矩阵【Leslie's Matrix】）。根据 1995 年意大利的生命率（其特征是低生育率和低死亡率）对 1995 年意大利的年龄分布（一个相对老龄化的社会）进行了预测，同时还预测了尼日利亚的（一个年轻的社会）。在一个中期时段（2030—2060）内，它们的年龄分布显得相当不同，因为，几个大队列的非常年轻的尼日利亚人步入了他们的生育期，即便以低的生育率也生育了若干相对大队列的儿童。经过较长的时期（2100—2200），年龄分布变得在统计上趋同，因为在出生人口上的较早的队列的连锁效应（ripple effect）已经消失。

尽管它们的年龄分布到 2200 年趋同，但是尼日利亚的整个人口增长将会更多。我们可以运用稳定人口模型来测量人口动量（population momentum），即当一个有着年轻的年龄分布的人口总体出现生育率下降时，发生的人口增长和暂时的人口学变化（Keyfitz，1985）。尽管尼日利亚的人口比意大利的人口会有更加实质性的增长，但是由于它过去最大的那些队列已经步入生育期，亦即当这些队列的年龄超过平均生育年龄时，所有这些增长都已经发生；而在给定如此低的生育率的条件下，新的队列只可能比老的队列规模小。

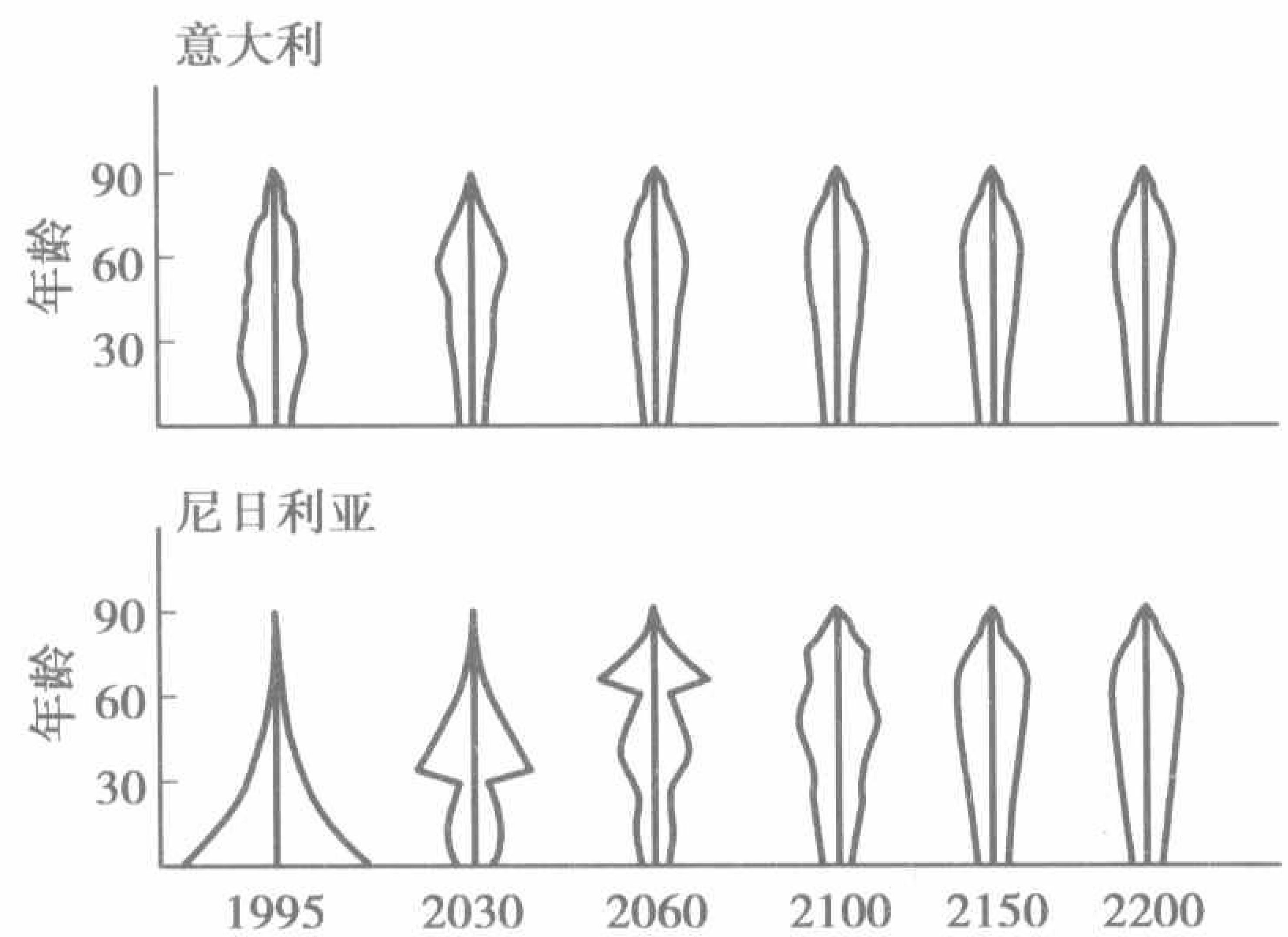


图 1 意大利和尼日利亚的相对年龄分布，均以 1995 年意大利的生命率作预测得出

资料来源：Preston，Heuveline，& Guillot（2001）；United Nations（1995，1996）

源于稳定人口模型简化了人口学会计法对于人口学变化的复杂的中期和长期的隐含意义有更加经验性的解释。具体而言，通过提高生育的平均年龄，由于被推迟的生育年龄对未来人口规模的实质影响（如更晚、更长、更少），计划生育策略已经受到了影响。由于相比于死亡率下降，生育率下降对人口老龄化、老年依存人口（与年轻成人相比的老年人口的相对规模）的更大影响，以及迅速的人口老龄化的周期性经济效应，政府的养老金计划也已经受到了影响。

更多的方法革新源于放松稳定人口模

型的主要假定。变量 r 法将稳定人口的属性推广到非稳定人口的属性,从而产生了更加广泛的人口预测和修正技术(Preston & Coale, 1982)。将稳定模型推广到具有恒定的、非零的年龄别迁移率的人口则促进了多国家和多地区稳定人口模型的发展(Rogers, 1995)。更多讨论和示例,参见普里斯顿的著作(Preston et al.,2001)。

——Randall Kuhn

(马妍译 赵锋校)

参考文献

Keyfitz, N. (1985). *Applied mathematical demography* (2nd ed.). New York: Wiley.

Lotka, A. J. (1998). *Analytical theory of biological populations: Part Ⅱ: Demographic analysis with particular application to human populations* (D. P. Smith & H. Rossert, Trans.). New York: Plenum. (Original work published 1939)

Preston, S. H., & Coale, A. J. (1982). Age structure, growth, attrition and accession: A new synthesis. *Population Index*, *48*(2), 217-259.

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and modeling population processes.* Oxford, UK: Blackwell.

Rogers, A. (1995). *Multiregional demography.* Chichester, UK: Wiley.

United Nations. (1995). *World population prospects: The 1994 revision.* New York: Author.

United Nations. (1996). *Demographic yearbook 1994.* New York: Author.

标准差(Standard Deviation)

一组数字和得分的主要特性之一是它的变异度或分散度【Dispersion】——各得分之间的差值有多大。标准差或许是一组得分变异度的最常用的量度。它之所以普及,一是因为它考虑了每一得分的确切数值,二是因为它的单位即得分的单位。此即,如果高度的测量单位是英寸,那么标准差的单位也用英寸表示。

通常,标准差还被视为平均数量度的一个补充,如算数均值【Mean】的补充。例如,许多城市在一年中有相同的平均气温,但是在不同季节中它们的气温变异度有很大的差异。在描述和比较这些城市的天气时,既比较它们的全年平均气温,同时也根据全年气温的变化程度,将各个城市分组比较,这样做能给出更丰富的信息。

同样,在研究中,为了对人们的行为作出恰当的推论,标准差可能与均值同样重要。例如,设想有一个项目,其目的在于提高学生测试成绩。当然,随着项目的开展,你期望看到平均成绩的提高。不过,这一提高可能以几种形式出现。如果项目以大致相等的程度帮助了每位学生,那么平均分将比项目实施前提高,但是标准差可能是相等的。如果项目只帮助了最好的学生,而其余的学生保持不变,那么你仍将获得均值的提高但标准差也随之增加(即最好的学生和最差的学生之间差距更大)。对于一些教育工作者而言,如果仅考查均值的证据,就不会认为后一个结果是他们所期待的。增加的标准差还给出了一个关于项目如何影响得分的重要线索。

某一整组或一总体得分的标准差 σ 可定义为:每一得分 X 与得分均值 μ 之差的平方之和除以得分的数目 N 的平方根,即

$$\sigma = \sqrt{\frac{\sum (X - \mu)^2}{N}}$$

之所以使用差值的平方是因为,若不然,正的差值和负的差值就会相互抵消。未取平方根时,量度 σ^2 即称为方差【Variance】,但它用单位的平方(如平方英寸)而不是单位本身来表示。在计算一个样本得分的标准差 s 时,即总体标准差的一个估计值,需要对公式进行微小的校正:

$$s = \sqrt{\frac{\sum (X - \overline{X})^2}{N - 1}}$$

其中,$\overline{X}$ 是样本得分的均值,N 被 $N-1$ 取代,这是对一个总偏差——基于样本数据量计算导致对总体标准差的低估——的校正。

同一方程的另一计算公式如下:

$$s = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N - 1}$$

为了说明这个公式,设想如下包含 10 个得分的样本:

8,26,15,9,13,20,21,10,15,12

对于上述得分,有 $\sum X = 149$;$(\sum X)^2 = 22\ 201$;$\sum X^2 = 2\ 525$;同时 $s = 5.82$。就这 10 个得分的样本而言,一个较好的统计汇总是它的均值为 14.90,标准差为 5.82。如果没有标准差,这一描述就会不完整。

——Irwin P. Levin

(马妍译　赵锋校)

标准误(Standard Error)

一个统计量的标准误是该统计量的抽样分布【Sampling Distribution】的标准差【Standard Deviation】,同时,它也被用来度量该统计量围绕它所估计的参数【Parameter】的随机变差【Random Variation】。例如,一个均值【Mean】的标准误是在给定样本量 n 的情况下,进行无限次抽样,所得各次样本的均值所生成的一个分布的标准差。标准误的大小部分地取决于样本量,同时基于大数定律,随着样本量的增加,标准误减小。此外,如果一个统计量是其参数的无偏估计量【Unbiased Estimator】,那么样本量越大,所估得的统计量越紧密地聚集在被估参数周围。

一个统计量的真实标准误一般情况下是未知的,因此标准误这个术语通常指对真实标准误的一个估计。虽然标准误可以由重复抽样获得的一个频次分布【Frequency Distribution】进行估计,但是,它们通常从一个单一样本中估得,并且代表了一个统计量的期望的标准差,就好像已经得到了大量的同类样本似的。从一个单一样本中估计标准误是可能的,因为某些统计量的抽样分布的性质已经是确定的。从一个单一样本估计标准误的方法展示如下,包括两个统计量的标准误:样本均值 $\overline{X}$ 和回归系数【Regression Coefficient】$\hat{\beta}_1$。

样本均值 $\overline{X}$ 的标准误

中心极限定理【Central Limit Theorem】确定,如果样本抽取自一个自身是正态分布的总体,那么即便样本量小,样本均值 $\overline{X}$ 的抽样分布也是正态分布的。更进一步,如果

样本量大,即使总体不是正态分布的,样本均值 $\bar{X}$ 的抽样分布也是正态分布的。因此,基于中心极限定理,几乎每一个样本均值 $\bar{X}$ 都被假定具有一个正态分布的抽样分布,并且因为 $\bar{X}$ 是总体均值 μ 的一个无偏估计量,抽样分布的均值即等于 μ,同时,标准误 $\sigma_{\bar{x}}$ 等于 X 的总体标准差 σ 除以样本量的平方根:

$$\sum_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

一般情况下,当总体标准差未知时,用样本的标准差 s 来估计标准误:

$$s_{\bar{X}} = \frac{x}{\sqrt{n}}$$

换句话说,$\sigma_{\bar{X}}$是 $\bar{X}$ 的标准误,$s_{\bar{X}}$是 $\bar{X}$ 的估计标准误。

回归系数$\hat{\beta}_1$的标准误

如果一个回归【Regression】模型中的因变量【Dependent Variable】(Y) 是一个抽取自独立且同一分布的正态分布的观测值的样本,那么参数的普通最小二乘【Ordinary Least Squares】估计即具有正态分布的抽样分布。在一个只有单个自变量【Independent Variable】的简单回归模型 $Y=\beta_0+\beta_1X+\varepsilon$ 中,β_1 的估计记作$\hat{\beta}_1$。$\hat{\beta}_1$的抽样分布呈正态分布,其均值等于参数 β_1,同时,估计的方差 $s^2_{\hat{\beta}_1}$等于随机误差【Random Error】项的估计方差 s^2 除以 X 的偏差的平方和。于是,$\hat{\beta}_1$的估计标准误等于 $s^2_{\hat{\beta}_1}$的平方根,

$$s_{\hat{\beta}_1} = \sqrt{\frac{s^2}{\sum(X-\bar{X})^2}}$$

公式中的分子项表明,误差项方差越小,$\hat{\beta}_1$ 的标准误越小;而分母项表明,较大的 X 的方差会得到一个较小的 $\hat{\beta}_1$ 的标准误,或者说一个更精确的估计量。

——Jani S. Little

(马妍译 赵锋校)

参考文献

Devore, J. L. (1991). *Probability and statistics for engineering and the sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Hamilton, L. C. (1990). *Modern data analysis: A first course in applied statistics*. Pacific Grove, CA: Brooks/Cole.

Kmenta, J. (1986). *Elements of econometrics* (2nd ed.). New York: Macmillan.

Sokal, R. R., & Rohlf, F. J. (1981). *Biometry* (2nd ed.). San Francisco: W. H. Freeman.

估计标准误(Standard Error of the Estimate)

大致来说,估计标准误(SEE)是根据估得的回归线计算因变量【Dependent Variable】(Y)的预测值所犯的平均"错误"。对于所有 X 的取值,假定 SEE 为常数。因此,当 $X=x_i$,$X=x_j$,$x_i \neq x_j$ 时,预测的 Y 值的平均误差是相等的。在经典的线性回归模型(classical linear regression model,CLRM)中,上述假定隐含在误差项方差为常数[$\mathrm{Var}(u_i|X_i)=\sigma^2$]的假设中。

SEE 和一个随机变量【Random Variable】的标准差【Standard Deviation】之间具有直观上的相似性。一个随机变量的标

准差只是该变量与其均值【Mean】的偏差的均方根。在一个回归框架中,这些偏差由扰动项【Disturbance Term】[$\hat{u}_i = (y_i - \hat{y}_i)$]代表。因此,SEE 可以被理解为扰动项的抽样分布【Sampling Distribution】的标准差,而该抽样分布根据假设以 0 为中心。SEE 也被称为回归的均方根误差【Mean Square Error】。

就一个有 k 个被估参数的回归而言,其中 k 是回归模型中自变量加常数项的数目,SEE 由以下公式给出:

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k}}$$

SEE 是估得的回归线对数据的拟合优度【Goodness of Fit】的一个量度。SEE 越小,回归线对数据的拟合越好。如果回归线完美地拟合了数据,那么数据中的每个观测值将会恰好落在回归线上,同时 SEE 为 0。一些研究者将 SEE 视为一个回归模型所偏好的拟合量度,认为该统计量具有许多优点。首先,它以因变量的单位为单位,可以在具有相同因变量的回归之间进行有意义的比较。同时,与另一个常用的拟合优度量度 R^2 类似,它不依赖于模型中自变量【Independent Variable】的方差【Variance】。

举一个例子来看如何计算 SEE。一位教师有兴趣预测学生在一门基础统计课程中的成绩。教师认为学生的最终考试成绩是如下自变量的线性【Linear】累加函数:(1)平均学分等级(grade point average, GPA);(2)学术能力评估测试(scholastic assessment test, SAT);(3)性别(gender);(4)在校年限(year)。使用有 15 个学生的随机样本【Random Sampling】,该教师得到了如下的预测方程【Prediction Equation】:

$$\hat{Y}_i = 9.14 + 1.78GPA + 0.11SAT - 2.67GENDER + 2.10YEAR$$

然后这些估计值被用于计算样本中每个学生的成绩(Y)的预测值(详见表 1)。

表 1 期末得分、预测得分和预测误差

序号	期末得分(Y_i)	$\hat{Y}$	$u_i = y_i - \hat{y}_i$	u_i^2
1	82	85.02	−3.02	9.11
2	90	85.43	4.57	20.90
3	78	81.39	−3.39	11.46
4	68	65.50	2.50	6.27
5	76	78.51	−2.51	6.630
6	69	74.93	−5.93	35.14
7	91	89.62	1.38	1.90
8	82	83.00	−1.00	1.00
9	91	88.69	2.31	5.33
10	66	66.69	−0.69	0.48
11	77	75.53	1.47	2.16
12	58	58.23	−0.23	0.05
13	75	69.64	5.36	28.78
14	85	82.64	2.36	5.59
15	74	77.20	−3.20	10.23
			$\sum_{i=1}^{n} u_i = 0$	$\sum_{i=1}^{n} u_i^2 = 144.69$

估得的回归线对观测数据拟合得如何?为了回答这个问题,教师运用表 1 中的数据和上文给出的公式计算了 SEE。

$$\hat{\sigma} = \sqrt{\frac{144.69}{15 - 5}} = \sqrt{14.47} = 3.80$$

平均地,模型将会产生 3.8 分的预测误差。回顾前文内容,一个小的标准误意味着一个较好的拟合(同时注意到 Y 的标准差是 9.68),我们可以得出结论:在这个例子中,估得的回归线较好地拟合了数据。

——Charles H. Ellis

(马妍译 赵锋校)

参考文献

Achen, C. H. (1982). *Interpreting and using regression*. Beverly Hills, CA: Sage.

Agresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Gujarati, D. N. (1995). *Basic econometrics* (3rd ed.). New York: McGraw-Hill.

标准分(Standard Scores)

标准分能使我们更容易地理解一个特定分数的含义。例如,如果知道一个学生在一项有 30 道题目的数学测验中答对了 20 道题目,这只能让我们对该学生的成绩有非常粗略的了解。若不了解同年级其他学生的成绩,我们就很难判断这个分数究竟有多好或者多不好。如果我们能获知同年级所有其他学生在同样的环境下参加该测验的平均得分是 17 分,那么我们可以知道得更多。至少我们可以知道这个学生的成绩“略微高于平均水平”。如果我们也能获知所有学生得分的标准差是 3 分,我们就能够更明确地知道该学生的成绩“高出平均值多少”,即高出均值 1 个标准差。假设我们已知同一学生在一项有 50 道题的口语能力测验中答对了 35 道题,同时测验的均值为 40,标准差为 5,那么我们可以知道这个学生在口语能力测验中的分数低于均值 1 个标准差。

正如上例所表明的,通过用整组的均值和标准差【Standard Deviation】对每个分数进行校准,从而标准化测验分数的集合,这样的做法起到以下两个重要功用:(1)它能让我们将一个学生的测验成绩与所有可比学生的成绩的整个集合联系在一起;(2)它能让我们比较同一个学生在一个测验中和另一个测验中的成绩。后一功用尤其能够用来判定一个学生的相对优势和劣势。

有多种方式可以做成标准分。例如,某些人格测量将结果校准至均值为 50 和标准差为 10,而有些智力测试则将评分校准至均值为 100 和标准差为 15。然而,社会科学家最常用的标准分的形式是 z 分数【z-Score】,即用分数 X 减去其均值 $\overline{X}$ 的差除以其标准差 s:$z=(X-\overline{X})/s$,来转换每个得分 X。

每一个转换得分都可以直接表示其高于或者低于均值多少个标准差。$z=+1$ 表示分数高于均值 1 个标准差;$z=-1$ 表示分数低于均值 1 个标准差;而 $z=0$ 表示该分数恰好精确地落在算术均值上,如此等等。若分数的原初分布形式已知,这些标准分还可以进一步转为百分位数【Percentile】。例如,在一项正态分布【Normal Distribution】中,只有 16%的分数高于均值 1 个标准差,因此一个值为+1 的 z 分数会落在第 84 位的

百分位数上(绝大部分统计书中都包含一张将正态分布的标准分转为百分位数的表)。

z 分数和其他用均值和标准差校准而得的标准分的另一个有用的特性是它们与测量单位无关。例如,一个人身高的标准分与其身高用英寸还是厘米测量无关。类似地,相关系数也是用标准分计算的,所以,我们就可以比较不同文化地区间受教育年限同收入的相关系数,而无须考虑不同国家的货币单位。暗含在标准分中的标准化具有很多实用的优点。

——Irwin P. Levin

(马妍译　赵锋校)

标准化回归系数(Standardized Regression Coefficients)

在最小二乘【Ordinary Least Squares, OLS】的多元回归【Multiple Regression】分析中,一个未标准化【Unstandardized】的回归系数 *b*,以变量的初始测量单位(如美元、千克、量表得分、年)描述了一个因变量【Dependent Variable】和一个自变量【Independent Variable】之间的关系。每一单位自变量的变动预计会产生 *b* 单位的因变量的变化。例如,受教育年限每增加 1 年可能引起年收入增长 $b=1\ 024$ 美元。

而一个标准化的系数 b^*,以标准差【Standard Deviation】为单位表达了同样的关系:受教育年限每增加 1 个标准差,可能引起年收入 $b^*=0.275$ 个标准差的增加。在 OLS 回归中,可以用两种方法计算标准化系数 b^*。一种方法是在回归【Regression】分析前,将所有的变量转换为 *z* 分数【*z*-Score】或者标准化变量【Standardized Variable】(就每个变量而言,用原值减去变量均值,再除以变量的标准差)。另一种方法是,就每一个自变量,用未标准化的系数乘以这个自变量的标准差,再除以因变量的标准差,即 $b^*=(b)(s_x)/(s_y)$,其中 *y* 是因变量,*x* 是自变量,s_y 是 *y* 的标准差,s_x 是 *x* 的标准差,而 *b* 和 b^* 分别是未标准化系数和标准化系数。

标准化系数常用于两个主要目的。首先,当自变量是用不同的单位进行测量时(例如,一个变量是受教育年限,另一个变量是 IQ 得分等),只根据未标准化系数很难判断两个或多个自变量之中哪一个对因变量有更大的影响。用不同单位测量(厘米、米或千米)的同一变量将会有一个更大的或者更小的未标准化回归系数。因为使用标准化系数将所有系数转化为一个共同的测量单位(标准差),所以标准化系数的单位是同样的,而无论变量最初的测量单位是什么。

如果一个自变量与第二个自变量相比具有更大的标准化系数,那么我们可以推论第一个自变量比第二个自变量有更大的影响,即使这两个自变量最初以不同的单位进行测量。其二,标准化系数也用于路径分析【Path Analysis】,去计算一个自变量对因变量的间接效应、总效应,以及直接效应。在一个完全递归【Recursive】的路径模型中,可能把自变量和因变量的相关系数“分解”为自变量的直接效应和间接效应(即自变量先影响其他自变量,后者转而影响应变量)。在只有一个自变量的特例中,标准化回归系数等于两个变量之间的相关系数。

戈德曼(Goodman, 1973a, 1973b)试图将标准化系数推到 OLS 多元回归之外,他用 Logit 模型【Logit Models】的未标准化系

数除以它的标准差,构建了他所谓的 Logit 模型的“标准化”系数,但是,这一系数不过是 Wald 统计量(常用于检验 Logistic 回归【Logistic Regression】系数的统计显著性【Statistical Significance】)的一个变体。戈德曼还试图用 Logit 模型,构建同路径分析相似的模型,但是他所提出的路径分析的类似模型并不允许计算自变量对因变量的间接效应或因变量同自变量间相关系数的分解(Fienberg, 1980)。

最近,梅纳德(Menard,1995)回顾了其他一些尝试构建标准化 Logistic 回归系数的研究,并且为完全标准化 Logistic 回归系数提出了一种方法,即区别于只将自变量进行标准化的部分标准化 Logistic 回归系数,将因变量和自变量都进行标准化。梅纳德(Menard,1995)和梅纳德所推进的工作表明,这种完全标准化 Logistic 的回归系数可以用解释标准化的 OLS 回归系数的方式予以解释,同时,它还可用于带有 Logistic 回归的路径分析,包括计算自变量对因变量的间接效应,以及自变量和因变量之间相关系数的分解。

——Scott Menard

(马妍译　赵锋校)

参考文献

Fienberg, S. E. (1980). *The analysis of cross-classified categorical data* (2nd ed.). Cambridge: MIT Press.

Goodman, L. A. (1973a). The analysis of multidimensional contingency tables when some variables are posterior to others: A modified path analysis approach. *Biometrika*, *60*, 179-192.

Goodman, L. A. (1973b). Causal analysis of data from panel studies and other kinds of surveys. *American Journal of Sociology*, *78*, 1135-1191.

Menard, S. (1995). *Applied logistic regression analysis*. Thousand Oaks, CA: Sage.

Menard, S. (in press). *Logistic regression*. Thousand Oaks, CA: Sage.

标准化测验(Standardized Test)

标准化测验常用于测量认知技能、能力、天资和成就,在心理学、教育学和其他社会学科领域有多种应用。标准化测量【Measure】也被发展用来评定人格、态度和其他非认知特征。然而,本条目的重点落在对认知特征的测量上,也就是通常所说的测验。

一项标准化测验是对一个清晰表述的构念【Consturct】的一个量具;具有清晰陈述的目的;通过使用高度结构化的程序发展而来;在严密控制的条件下予以实施;具有一个用于得分解释的系统基础;还拥有不同的技术特性。

构念定义

标准化测验是对构念的测量。AERA/APA/NCME(1999)的测验标准(test standards)把构念界定为“广义上是一项设计的测验所要测量的概念或特征”(p. 5)。在天资测验中,测验可能被构建来测量构念——空间能力。在成就测验中,测验可能被发展来测量构念——数学成绩。标准化测验的主要组成部分是构念定义,即对所要测量的构念的一种清晰陈述。构念定义为发展待测构念的测验和评估测验的效度提供了基础。

测验的发展

系统化的程序被用来发展标准化的测验。就一项测验如何评估构念,测验的发展程序帮助提供一个操作化的定义。对于标准化的天资测验,包含在测验中的不同题项类型被具体指明,同时也具体指明了用以制作题项的规则。对于标准化的成就测验,要制作细节化的详述说明表格,以指出包含的题项类型、内容领域和题项的认知技能的水平。这类测验的详述说明为发展测验的人提供了一个用以测验构念的蓝图。

标准化测验也有统计上的详细说明。在测验题项被纳入测验的操作版本之前,前测程序常用来理解测验题项对应试者起多大作用。题项难度和题项区分度的统计指数通常根据前测来计算。统计上的详细说明表明,测验包含了多少不同难度和区分度水平的题项。

具体的内容和统计上的详述说明是发展备选测验的前提条件,而后者是不同的测验题项的集合——可以期望它们能以非常相似的意义和统计属性生成测验得分。备选形式强化了测验的安全性,容许对应试者实施多次测试。

测验的实施

标准化测验有严格控制的实施条件:应试者被给予清楚的测验说明;以相同的方式给予所有应试者测验手册或其他测验材料;测验的时间也要仔细控制;所有应试者要使用同样的回答记录工具或程序;还要控制和指定其他工具的使用,如计算器和草稿纸。

应试者的回答

一些标准化测验使用多个回答选项的样式,即要求应试者从一组可能的答案中指出正确的一个。也有其他的客观评分样式,如对错题。在一些测验中,可能要求应试者对问题给予简短的答案或扩展的答案,如一篇短文测验;而在另一些测验中,要观测应试者的行为。

评分程序

在标准化测验中,常使用清楚定义的评分程序。就客观化的评分测验而言,原始分通常被定义为应试者正确回答的题项的数量。另外,有时也使用更复杂的原始分数,如部分信用评分。就那些应试者要给出简短或扩展答案的测验而言,则要使用人类判断来为回答评分。在这种情况下,要使用具体的评分说明,即对于回答评分的判断给出清晰的指导。

分数量尺

标准化测验的原始分被转换为量尺【Scale】分数以便于分数的解释。分数换算可能是规范的。例如,针对智力测验实施规范的研究,其中一群据认为能代表一个特定总体的应试者(如美国的14周岁的儿童)被施以测验。从规范的研究中得到的原始分数的均值被设定为一个100的量尺分数。其他的规范点被用于设定分数量尺上的其他点。

熟练水平评分有时被用于教育成就测验。在实施于一个标准场景的熟练水平评分中,评分人将根据应试者在某个熟练水平上,还是高于特定水平来确定他们的原始得分。一组经常使用的熟练水平是基本(basic)、熟练(proficient)和高级(advanced)。

规范和熟练水平为测验的使用者提供了解释分数的基础。就其本身而言,它们是标准化测验的主要组成部分。

测验的保护

测验的分数代表要测量的构念是很重要的。然而,如果测验的问题让未来应试者知道了,他们的表现可能不能反映应试者对那个构念的熟识程度。由于这一原因,许多标准化测验项目都有详细制订的保护程序。这些程序包括使用备选形式、严格控制测验材料,有时在测验实施完成后,进行统计分析来发现保护中的不端行为。

信度

为了作出重要决策,标准化测验的制作要能产生有足够信度【Reliability】的分数。关于测验时间,要对测验包括的问题的数目、问题的类型作出决策,以确保分数对于它们的预期目的有充分的信度。

效度

测验的效度【Validity】是证据在多大程度上支持对测验分数的解释。一项测验的效度评估过程包括收集那些能够为分数的解释提供一种似乎科学的基础证据。效度评估始于对被测构念的陈述、测验的目的,以及分数构成的解释。要收集不同类型的证据来支持测验的效度。效度证据的来源可以以测验的内容、应试者回答的过程,以及对测验的内在心理测量学结构的一项研究为基础,也可以以测验分数同其他变量的关联为基础。那些说明测验没有被外在于所测构念的变量影响的证据是效度评估过程的一个重要组成部分。合理的效度评估过程要整合不同来源的证据以支持特定使用目的下测验分数的解释。

特别议题

近年来,一些标准化测验的实施已经转移到计算机上,从而对发展标准化测验造成新的挑战。心理测量学【Psychometric】的测验理论,如项目反应理论【Item Response Theory】和概化理论【Generalizability Theory】,近年来都有重要的发展,推进了对标准化测验的技术特性的更好理解,同时,发展了用于实施和评分标准化测验的新程序。

——Michael J. Kolen

(王玥译　赵锋校)

参考文献

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Anastasi, A., & Urbina, S. (1996). *Psychological testing* (7th ed.). New York: Prentice Hall.

Linn, R. L. (Ed.). (1989). *Educational measurement* (3rd ed.). New York: Macmillan.

Plake, B. S., & Impara, J. C. (Eds.). (2001). *The fourteenth mental measurements yearbook.* Lincoln, NE: Buros Institute of Mental Measurements.

标准化变量(Standardized Variable)

参见 z 分数【*z*-Score】、标准分【Standard Scores】。

认识论立场(Standpoint Epistemology)

认识论立场是从女性主义对社会科学中女性地位的缺失或女性的边缘地位的批判中发展而来的。尽管社会科学按说是客观的【Objectivity】和价值中立的,但是女性主义者认为社会科学大部分是从男性的视角和男性的利益出发的。为了让女性的生命意义更加可见,必须从她们的视角来分析。认识论立场的重要性在于,它挑战了那些根据普遍论的男性假定而作出的对社会生活的描述和分类。关键的理念是,女性和男性的生活有着不同的轮廓,创造不同种类的知识。女性的压迫经验产生特定且具有特权的理解。对不同知识的接近能够揭示其他形式人类关系的存在,而从占据支配的男性位置看,这点是不可见的。因此,认识论立场为新的且更可信的对性别权力和关系的洞见提供了可能性。这对于女性主义研究【Feminist Research】的实践非常重要,例如,强调倾听女性声音的重要性。这也为扩展对其他群体(如少数族群、民族和残障群体)的理解提供了可能性。

相关理念的差异存在于,一个女性主义的立场应当如何被思考,以及它对知识的隐含意义,如有些人使用词语理论(theory),而不使用认识论(epistemology),因为后者是同确定性及中立性连在一起的。也有人提出,立场是一个实践的成果,而不是一种归属的状态;被支配本身并不带来对于不平等和压迫的觉醒。立场方法的一个主要困难是,存在着用概化的女性视角取代概化的男性视角的风险。近来,女性主义者质疑一种观点,就是女性的本质是基础性的或是固有的吗?女性的压迫是普遍的、同质的、共享的吗?她们转向强调女性之间的差异和不同,特别是聚焦于那些源于民族、性别、残障、阶级和年龄的结构地位产生的差异。这导致不同立场的激增,例如,一些黑人女性主义者和同性恋女性主义者的立场。然而这一方法也产生出多种不可比较的立场和相对主义【Relativism】立场的共存问题。它假定,一种结构位置,而不是多种位置——这可能引起矛盾或紧张——只能提供特定的知识。还存在另一种风险。此即围绕特定立场发展出正统学说,以及存在和观察的方式,据此就可能判定某些女性偏离了既定规范。

认识论立场的有效性和意义已经被广泛地讨论,分歧也比比皆是。然而,存在着一些共识,即它突出了女性的经验同女性主义知识之间的关系,以及知识同权力之间的关系。不存在同利益关系无涉的知识,而正是立场的采纳将研究者的关注投向位于知识生产中心的主题。

——Mary Maynard

(王玥译　赵锋校)

参考文献

Collins, P. H. (1997). Comment on Heckman's "Truth and method: Feminist standpoint theory revisited": Where's the power? *Signs*, *22*(21), 375-381.

Hartsock, N. C. M. (1998). *The feminist standpoint revisited and other essays*. Boulder, CO: Westview.

Smith, D. (1997). Comment on Heckman's "Truth and method: Feminist standpoint theory revisited." *Signs*, *22*(21), 392-397.

Stata

Stata是一种以通用为目的的统计软件。该软件近年来在社会科学领域获得了广泛应用,其部分原因在于它易于执行许多由用户自己写作,且便于维护的宏。如欲获取更多信息,请参见该软件的相关资料。

——Tim Futing Liao

(赵锋译校)

统计量(Statistic)

一个统计量只是样本数据的数值汇总。数据分析者经常使用,又为大家所熟知的样本统计量,包括均值、样本方差【Variance】和标准差【Standard Deviation】,样本百分数【Percentile】和中位数【Median】及此类数值。这些数值综括了样本的定量数据。就类别【Categorical】数据样本而言,样本比例、百分比和比率都是典型的样本统计量,可用于汇总数据信息。换言之,一个统计量只是从样本数据中计算出的综括的量。

就整个调查总体【Population】而言,统计量的对应部分被称作参数【Parameter】。一个统计量至少在以下两个方面区别于总体参数。第一,一个统计量描述或汇总样本的特征,而参数是指从中抽取该样本的总体的数值汇总。请注意,样本统计量的值依赖于抽取到的样本。由于样本统计量的值在样本间变化,样本统计量有时被称作随机变量【Random Variable】,以此反映它们的值随抽取到的样本而变化的特性。

第二,参数值通常是未知且不易获得的量,而统计量可以由样本数据直接计算得到。大多数社会研究的首要兴趣是总体的参数,而非从选取的观测值中计算得到的样本统计量。因此,如果某个研究所感兴趣的总体很小且数据来自整个总体,那么研究者与其考察样本的统计量,不如直接考察总体参数。

不过,在大部分社会研究中,研究者需要使用样本统计量来推论总体的参数值,有如下原因:(a)社会研究的总体通常非常大,以致没有实际可能收集整个总体的数据;(b)如果使用了恰当的抽样技术和统计方法,研究者在推论(或猜测)总体参数的真实值时可以获得很好的精度。例如,由专门的民意调查公司和研究机构,比如盖洛普(Gallup),一般社会调查(General Social Survey)和国家选举研究(National Election Study)所抽取的样本,通常包括1 000~2 000个被调查主体。

用一个样本统计量去估计或推论参数值的一个主要问题是该估计精确到何种程度。换言之,数据分析者需要知道用样本统计量来预测【Prediction】参数值的可能的精确度。传统用以表示预测的不确定程度的方法是报告参数点估计的置信区间【Confidence Interval】。人们通常令样本统计量等于总体参数的点估计【Point Estimates】(此即,样本均值等于总体均值,样本方差等于总体方差)。此外,研究者在数据收集之前还需设定估计的“欲得精度(desired precision)”。这一欲得精度(也被定义作误差界限)转而可以帮助研究者确定获取可信样本统计量所需要的样本量。关于样本统计量的置信区间,以及样本量与使用样本统计量估计参数精度之间

的关系问题，读者可以参见威廉·海斯(Hays，1994)，艾伦·阿格雷斯蒂和芭芭拉·芬利(Agresti & Finlay，1997)，以及理查德·约翰逊和古里·巴塔查里亚(Johnson & Bhattacharyya，2001)等人的书籍。

——Cheng-Lung Wang

(赵锋译校)

参考文献

Agresti，A.，& Finlay，B.(1997). *Statistical methods for the social sciences*. Englewood Cliffs，NJ：Prentice Hall.

Hays，W. L.(1994). *Statistics*(5th ed.). Fort Worth，TX：Harcourt Brace.

Johnson，R. A.，& Bhattacharyya，G. K.(2001). *Statistics：Principles and methods*(4th ed.). New York：Wiley.

统计比较(Statistical Comparison)

大多数社会科学家会同意社会科学的目的之一是比较，而比较可以给予我们通过其他方法无法得到的洞见。比较的传统可以追溯至社会科学的起源，在涂尔干和韦伯的著作中都可以找到相应的证据。目前，具有比较性质的社会研究涵盖了从社会和民族-国家的比较研究【Comparative Research】，到使用比较方法【Comparative Method】的小样本量 n 的个案研究。只要比较研究者使用某种形式的数值数据，他们通常会使用某种统计比较。在因特网上搜索“统计比较(statistical comparsion)”就会找到成百上千条包含此术语的线索，其中有些属于社会科学研究。但是，研究需要做哪些类型的统计比较呢？其中，有些只是两组数字的列表比较，比如来自两群人的比率和平均数，另一些可能包括某些形式的显著性检验【Significance Testing】，其目的在于提供样本【Sample】统计量所来自总体【Population】的推论【Inference】。本词条的目的在于概要式地介绍众多可能的统计比较。

均值的比较

通常，研究者有兴趣了解是否一群人比另一群人在某个更高的水平上具有某些属性。这就是两个均值【Mean】的比较。例如，人们可能有兴趣比较 a 和 b 两个群体的收入均值[$E(Y_a)$ 和 $E(Y_b)$]。虚无假设【Null Hypothesis】陈述 $E(Y_a)=E(Y_b)$，同时，当分析的是样本数据时，通常会应用t 检验【t-Test】。此类比较是实验【Experiment】的核心。想要了解与此相关的更多内容，可以参见安德森等人(Anderson，1980)的著作。

尽管在一项实验中，选取主体进入处理【Treatment】组和控制组【Control Group】是随机的，但是，在多数社会科学研究中，比较中的组变量是固定的(比如性别、人种和国别)，无论个体是否抽取自同一样本或不同的样本。

两组均值比较的观念可以扩展到多组比较【Multiple Comparisons】。有多种方法可以比较多组均值，比如 Bonferroni 技术【Bonferroni Technique】、Scheffe 检验【Scheffé'e Test】，以及 Tukey 法。单因子方

差分析【Analysis of Variance】的一个主要目的也是用以估计来自多个组别的回答均值是否存在差异。

方差的比较

如果人们有兴趣估计两个分布在方差上的差异,比如收入,人们就是考虑两个方差的等值性;或者换一种表述形式,即虚无假设是 $\text{var}(Y_a)=\text{var}(Y_b)$。$F$ 检验【F Test】可以用于方差相等的估计。这是一项有用的检验,特别是当两个收入均值可能一致,但是一个收入分布可能代表一个完美的中产阶级社会,其中每个人的收入水平都位于均值附近,而另一个分布来自一个非常不平等的社会,其中收入能力相差很大。想要分析多个群组间的方差不等性,Levene 检验【Levene's Test】是适当的选择。

分布的比较

尽管方差是分布的一项属性,但是还存在着几种正式的比较分布的方法,即可以通过研究它们的比例分布、相对分布和理论分布来比较分布。上述所有的比较分布的方法考虑的都是两个(或多个)分布之间形状和一般的差异,完全不同于两个(或多个)参数间的值的差异,比如均值和方差。

应用洛伦兹曲线和基尼系数【Gini Coefficient】的方法显示了第一种方法的应用,其中研究者询问关于变量 Y(或收入)的累积比例,即由累积的总体份额所显现出的比例。这一方法通常用于研究收入不平等,而两个(或多个)交叠的洛伦兹曲线可以直接比较两个(或多个)分布,同时,与之相应的基尼系数可用于估计比较的程度。

基于所谓的相对分布,相对分布法【Relative Distribution Method】考察的是整个分布的比较。相对分布可以简单地定义为通过包含两个被考察的分布的计算而得到的一个分布。可以研究相对分布的累积分布函数或概率密度函数【Probability Density Function,PDF】。例如,两个收入分布的相对分布的 PDF 是均匀的。当然也可以针对某个理论分布来比较分布。例如,研究者可能有兴趣知道是否两个分布都服从正态分布【Normal Distribution】。存在着这样的可能,即来自一个样本的分布服从正态分布,而来自另一个样本的分布服从对数-正态分布。在判别一个类似广义线性模型【General Linear Model】的统计假设【Assumption】是否可能被违背时,可能同样要进行这类比较。

比率的比较

有时研究者的兴趣在于比较事件发生的比率,而不是某个变量的均值。例如,一项研究可能想比较两个总体的失业率。因为存在着潜在的混杂【Confounding】因子的效应,所以,不经过某种调整【Adjustment】,两个比率可能不能直接相比。这时,一个可选的方法是比率的标准化【Rate Standardization】,即与一个标准的总体构成相比,来标准化比率(也就是,清洗混杂效应的比率)。

回归参数的比较

当执行一项多元回归分析【Multiple Regression Analysi】,以及研究某些自变量【Independent Variable】对因变量【Dependent Variable】的影响时,研究者常常有兴趣比较不同样本或群组间的相关参数。例如,在研究收入时,人们可能想要比较男性(m)和女

性间(f),在年龄(β_1)和受教育程度(β_2)的效应上的差异。于是可能有这样三项虚无假设:

1. H_0 : $\beta_{1m}=\beta_{1f}$;
2. H_0 : $\beta_{2m}=\beta_{2f}$;
3. H_0 : $\beta_{1m}=\beta_{1f}$和$\beta_{2m}=\beta_{2f}$。

这里有几种方法可以检验假设1和假设2。一个通行的方法是基于合并的样本,计算一个带有两个交互项的回归分析,其中一个是年龄和性别哑变量【Dummy Variable】间的交互项,而另一个是受教育程度和性别哑变量间的交互项。对两个交互项之一做t检验可以估计假设1或假设2。用其他方法则可以估计假设3。一种方法是计算两个回归模型,其中一个带有两个交互项,而另一个没有交互项;然后,进行R方【R-Squared】增量的F检验。假设3的一个变化是不仅仅检验年龄和教育的参数的等值性,还同时检验性别哑变量的参数的等值性。估计的方法是Chow检验【Chow Test】,即评估两个性别组的收入水平是否完全不同。

广义线性模型的比较

上述讨论的假设可以容易扩展至其他回归类型的模型,比如Logit【Logit】、Probit【Probit】、多项Logit【Multinomial Logit】,以及Poisson回归【Poisson Regression】模型,而它们都是广义线性模型的成员。尽管假设的陈述可能很相似,但是检验会很不同。有兴趣的读者可以参看廖福挺(Liao,2002)的著作。

其他统计模型的比较

相似地,人们还可以用更高级的模型,比如潜类分析【Latent Class Analysis】、结构方程模型【Structural Equation Modeling】,以及多层次分析【Multilevel Analysis】来进行比较。一个人的赚钱能力可能更适于当作一个潜变量【Latent Variable】来研究。属于同一较高水平单位的个体有更多相似的地方,而他们的行为方式可以被当作因变的观测值【Dependent Observations】,这时,社会结构的多水平性质不仅可以而且也应当被纳入考量。这类方法,以及其他类似方法的统计比较在廖福挺(Liao,2002)的著作中有非常详尽的讨论。

——Tim Futing Liao

(赵锋译校)

参考文献

Anderson, S., Auquier, A., Hauck, W. W., Oakes, D., Vandaele, W., & Weisberg, H. I. (1980). *Statistical methods for comparative studies: Techniques for bias reduction*. New York: Wiley.

Liao, T. F. (2002). *Statistical group comparison*. New York: Wiley.

统计控制(Statistical Control)

统计控制是指在多元分析【Multivariate Analysis】中将某个特定的自变量【Independent Variable】对因变量【Dependent Variable】的效应从其余自变量对因变量的效应中分离出来的技术。在统计分析中,此项技术对于作出有效推论【Inference】极为

重要。因为人们只有保持其他变量不变才能将某个具体的效应归之于某一自变量。在统计文献中，语汇，如“保持不变”“控制住”“解释了”或“纠正了……的影响”都用于指统计控制技术，常常可以互换使用。统计控制能帮助我们更好地理解某一自变量的效应。例如，在多元分析中，由于可能存在混杂【Confounding】变量或压抑【Suppressing】变量 $X_j(j>1)$ 对自变量 Y 的影响，X_1 对因变量的影响就可以通过保持其他变量恒定而分离出来。

当我们在估计一个特定模型【Model】时，技术上如何实现统计控制呢？假设我们发现竞选开销（X_1）增加特定选区层面上的投票率（Y）。为了估计竞选开销效应，如果我们将作为备择解释的控制变量引入我们的模型，我们就会知道这一关系是否被其他变量所混杂、压抑或者就是虚假【Spurious】关系。比如，某个选区族群被期望的封闭程度（X_2）就可能是一个备择解释。预计封闭的族群尤其会动员投票者在选举日参加投票。考虑下面的线性回归模型，即

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\varepsilon$$

竞选开销对投票率的独特效应 β_1 在经典回归【Regression】框架下可被估作

$$\hat{\beta}_1=\frac{\sum(X_1-\hat{X}_1)(Y-\hat{Y})}{\sum(X_1-\hat{X}_1)^2}$$

其中 $\hat{X}_1=\alpha_0+\alpha_1X_2$ 表示 X_1 对 X_2 的回归预测；$\hat{Y}=\gamma_1+\gamma_2X_2$ 表示 Y 对 X_2 的回归预测。由此，竞选开销的效应 $\hat{\beta}_1$ 仅依赖于 X_1 同 Y 的同时变化。因为 X_2 在计算项 $X_1-\hat{X}_1$ 和 $Y-\hat{Y}$的线性影响已经被减去了。结果是 $\hat{\beta}_1$ 所据以计算的项与 X_2 无关。在估计 X_1 对 Y 的特殊效应时，使得任何可能归之于 X_2 的可能干扰保持恒定，以致 X_1 和 Y 的关系不再依赖于 X_2。统计控制通过纠正 X_2 的影响分离出 X_1 对 Y 的独立效应 $\hat{\beta}_1$。同时，统计控制也解除了 X_1 的所有影响，分离出 X_2 对 Y 的独立效应 $\hat{\beta}_2$。这就是说，这项技术在处理关键的解释变量与其他（控制）变量时并无区别。统计控制背后的逻辑可以应用到更多的自变量上，甚而可以用在广义线性模型【Generalized Linear Models】上。

然而，统计控制伴随着假定【Assumptions】的简化。理想条件下，研究者希望内在地保持混杂因素不变，比如通过随机化实现的实验控制。但是对于观测数据而言这是不可能的，研究者在统计控制之前，必须清晰地指明混杂因素。而且，统计控制假定可能的混杂效应是线性且可累加的，此即，它们独立于模型中任何其他变量的特定水平。回到前面的例子，不管所期望的族群封闭程度到底有多大，它对竞选开销和投票率的影响都是相同的。如果这些假定被证明是错误的，研究者可以将自变量的乘积项带入模型，以此控制可能的条件效应、交互效应【Interaction Effects】，或在某个子总体上假定成立的情况下，针对该子总体获得模型的约束估计量。

——Thomas Gschwend

（赵锋译校）

参考文献

Gujarati, D. N. (2003). *Basic econometrics* (4th ed.). Boston: McGraw-Hill.

Lewis-Beck, M. S. (1995). *Data analysis: An introduction*. Thousand Oaks, CA: Sage.

Mosteller, F., & Tukey, J. W. (1977). *Data analysis and regression*. Reading, MA: Addison-Wesley.

统计推论(Statistical Inference)

统计推论是当我们仅有样本【Sample】统计量时,对总体参数作出概率【Probability】陈述的过程。总体参数是难以确知的固定量。假如我们知道总体,我们可以计算(而非估计)总体参数。然而,我们通常不知道总体的情况,因此我们必须从样本估计总体参数。由于存在着抽样误差【Sampling Error】,此类参数估计【Parameter Estimation】总是伴随着某种程度的不确定性。由此,我们对反映总体参数的可能性作出统计推论。

理解统计推论需要理解抽样分布【Sampling Distribution】的概念。抽样分布是从某总体中抽取所有可能的样本量为 N 的随机样本计算得到的样本统计量的概率分布。通常,抽样分布由两维的图表示,横轴上排列了样本统计量的所有可能的值,纵轴上给出了每个样本统计量的可能值的概率。现实中,社会科学研究者一般仅从抽样分布上抽取单一样本。然而,抽样分布的知识使得我们能够对与真实总体参数关联的样本统计量作出概率陈述。

我们依赖统计学的基本定理来对某个样本统计量是否反映真实的总体参数作出概率陈述。例如,统计推论的一项相当弱的工具是 Chebychev 定理。该定理称某个估值落于其抽样分布均值 k 个标准离差之外的概率是 $1/k^2$。比如,设若我们从任意形状的分布上随机抽样,并得到某总体的一个点估计【Point Estimate】。那么,Chebychev 定理称这一估值落于被估参数的抽样分布的真实均值两个标准差【Standard Deviation】之外的概率仅有 1/4。

作出统计推论的一组更有用的定理集合是中心极限定理族。中心极限定理【Central Limit Theorem】的 Lindberg-Levy 变式称,若样本量 N 的重复样本取自均值为 μ,方差为 σ^2 的任意形状的概率分布,那么随着 N 变大,样本均值的抽样分布逐渐接近总体均值为 μ,其方差等于 $\sigma^2/\sqrt{N}$ 的正态分布。这一结果非常重要,因为该定理说,无论总体分布有何种分布形态,其样本均值的抽样分布都将呈正态分布,且抽样分布的均值就等于真实的总体参数值,抽样分布的方差也可由总体方差和样本量 N 计算得出。根据此信息,研究者可根据样本统计量和它的方差作出总体参数的概率陈述。

还有许多其他的分布可用于统计推论。不过,它们依赖于所要推论的统计量的性质和分析者作出的预设。例如,将总体方差视作已知可能是不现实的。在此情形下,研究者经常使用 t 分布。其他广泛使用的分布有卡方分布和 F 分布。卡方分布常适用的情形是该统计量包含多个呈正态分布的变量的平方和。F 分布常用以估计两个独立的卡方变量的比值。

——B. Dan Wood
Sung Ho Park
(赵锋译校)

参考文献

Blalock, H. M., Jr. (1979). *Social statistics*. New York: McGraw-Hill.

Casella, G., & Berger, R. L. (2001). *Statistical inference* (2nd ed.). New York: Wadsworth.

Judge, G. G., Hill, R. C., Griffiths, W. E., Lutkepohl, H., & Lee, T. (1988). *Introduction to the theory and practice of econometrics*. New York: Wiley.

Spiegel, M. R., Schiller, J., & Srinivasan, R. A. (2000). *Schaum's outline of probability and statistics* (2nd ed.).New York: McGraw-Hill.

统计交互(Statistical Interaction)

对数理统计学家而言,统计交互指下列情形,即在多元分析【Multivariate Analysis】中变量间的关系随变量的取值不同而变化。在最简单的三变量情形中,两个变量 *A* 和 *B* 的关系根据第三个变量 *C* 的取值而变。假如不存在交互作用,变量间的关系是可累加的——就单个因变量的情形而言,此即意味着,我们可以总和自变量的效应,从而得出所有自变量对因变量的总效应。这一总效应是恒定的,不随各自变量的取值发生变化。交互效应表示,统计关系是非加性【Nonadditive】。物理学家将可累加的效应视作叠加(superposition)。如果存在交互效应,那么叠加瓦解。此类叠加失败是非线性【Nonlinear】系统的一个特征。

有多种统计程序可用于探测交互效应。方差分析【Analysis of Variance,ANOVA】技术常允许有包含交互项的模型设定。此外,还有同方差分析技术类似的探索性统计分析技术。在报告交互效应时有两类变化的样式:一种是将交互项写到线性方程中;另一种是报告模型的决定要素,而决定要素设定交互效应是显著的。例如,在研究住房和健康关系时,研究发现所有年龄低于 65 岁的群体,住房质量和健康质量的关系都是正向的。对那些高于 65 岁的人,这一关系是负相关。其因果过程是,不健康的状态是给老年家庭分配非常高质量居住条件的决定要素。一般而言,住房条件是健康状况的因。而对老年人而言,健康则是住房条件的因(Byrne, McCarthy, Harrison, & Keithley, 1986)。然而,此类直接的解释并非总是可行的。

马什(Marsh, 1982)指出"从技术的观点看交互效应是令人头痛的,但是从实际的社会学理论出发,交互效应则是激动人心的"(p.91)。从实在论【Realism】视角看,交互效应的存在挑战了指导传统统计推理的整个分析方案。如果我们用变量【Variable】的语言,正是交互效应的存在证明因果过程是局部和条件的,而非普遍的。换言之,不可能指明一般规律,因为"情况[交互变量的值]随条件而变"。另一种说法是,我们需要处理复杂和偶然的生成机制。此外,交互效应同样可能意味着突生(emergence)的存在——概述此思想的科学表达是"总体大于部分之和"。如果我们处理的是带有突生性质的复杂系统,那么其中一项论争是,变量的效应是具体的而非现实的,因此,我们需要陈述整个系统的特性,而不是仅报告系统的人为的非现实的抽象属性。至少,交互效应的存在提醒我们复杂关系的可能性,而且我们可以论证,在多元分析中对交互效应的探测使得任何用"线性(linear)"方法对所研究系统的因果关系的推理变得无效。

——David Byrne

(赵锋译校)

参考文献

Byrne, D. (2002). *Interpreting quantitative data*. London: Sage.

Byrne, D. S., McCarthy, P., Harrison, S., & Keithley, J. (1986). *Housing and health*. Aldershot, UK: Gower.

Marsh, C. (1982). *The survey method*. London: Allen and Unwin.

统计软件(Statistical Package)

统计软件是所有旨在帮助统计分析和数据探索的软件的集合。定量和统计分析中的主要工作需要统计软件完成。要想正确、有效地使用定量统计方法,就必须掌握统计软件的使用。尽管研究者有时可以使用通用的汇编语言,如FORTRAN,C++还有Java来执行统计分析,然而统计软件比起此类的"人工编码"有许多潜在优势。第一,统计软件可以节省研究者的时间和精力:统计软件提供了一组分析工具,而这样一组分析工具可能会耗费单个程序员几年的功夫去编写。第二,统计软件能够赋予数据管理、数据可视化和统计分析统一的操作框架和共同的交互模式。最好的情况下,统计软件中的工具可以被方便地组合和扩展,从而研究者可以在先前研究工作的基础上发展新的工具。此外,即便统计程序简单得足以实现手工编写或者由于太专门而需要手工编写,统计软件所提供的交互模式可以让其他使用者更容易地使用新程序。第三,统计软件比起汇编程序生成正确结果的可能性更大。即使是简单的任务,如计算标准差,若想通过编写生成可靠而精确的结果,也会相当棘手。统计软件可以就其中的程序缺陷和数值问题进行反复检测,同时现代数值方法的研究者比起通常的研究者更熟悉软件的编写工作。第四,统计软件的运行可能要快于一般的编写运行:在编写工作中,相当的精力和改进工作投入为某个特殊分析或统计工具选择算法上,以及为更好地适应具体的计算机架构而改写软件上。

当研究者使用统计软件时,他面临的主要问题是在几百种软件中选出合适的软件,从而确保用软件得到的结果既正确又可以复得。显然,尽管在多数大型软件中有相当部分的重合,然而每一软件根据它所支持的性能而相互区别。商业巨人,如SAS,SPSS以及大型开源软件"R",提供了几千种用于数据管理、数据可视化和统计分析的工具。这些软件的交互模式相当完善:从SPSS的"点-击风格(point-and-click style)"到SAS和R的基于命令行的交互。在不必坚持某种软件的专属性能的前提下,研究者可以考虑由现代统计语言"S"所写成的软件。这类软件功能全面,既有商业软件"S+",也有开源软件"R"(作为入门参考,读者可参见Venables & Ripley, 2002)。

其次,软件在速度、精确性和扩展性上也显著地相互区别。当研究者使用某个统计软件时,他应当提出这样的问题:该软件检验过其精确性吗?它是否提供了相关统计分析的算法以及其他研究者感兴趣的量的说明文件?它是否可以有效地处理拟议的研究方案所要求的数据量和数据结构?软件中有哪些编程工具可以使用?编程语言是否支持好的编程实践,如面向对象设计的编程模式?软件的内部程序可以审阅和扩写吗?软件使用外部程序库的易用程度如何,或者软件是否易于作为更大的程序编写和数据分析环境的组成部分?

尽管统计软件被视为统计分析的必要工具,然而人们常认为它们是可以高度互换的。由此,研究者一旦得到了统计结果,便常常疏于在之后的出版物中报告所使用的软件。实际上,统计软件在精度和可靠度上有显著不同,因此,报告的统计结果可能取决于特定的软件和版本,以及分析时所用到的计算选项(McCullough & Vinod, 1997)。

有几种方法可以确保精确的结果。研究者可以试着用不同的软件和不同的算法,看能否得到重复的结果。当软件支持额外的精度选择时,分析时增加计算的数值精度

可用于避免某些特定类型的数值误差。另一种方法是敏感度分析。其中,可将少量的干扰引入计算、数据和(或)初始值中,以此检验所得到的解在数值上是否稳健(Altman, Gill, & McDonald, 2003)。

最后,应采取以下三个步骤来增加统计结果的可重复性。第一,因为复杂分析可能依赖特定的计算策略(可以是研究者选定的也可以是内在于统计软件中的),所以报告(或保存)用于执行分析的编码、软件名称和软件版本以及运行时的所有计算选项是确保重复结果的必要条件。在某些情况下,当确切的重复极重要的时候,保存统计软件的拷贝可能需要获得授权;尽管一些软件可以重复旧版本的编程句法和计算选项,然而多数软件随时间的推移不仅会改变其编程句法,还会改变其编写方式。第二,许多社会科学数据组难以或不可能重新生成——在此情况下,公开地存档所有用于分析的唯一数据常是成功重复的必要条件。第三,统计软件使用的数据格式常常是专门的且易随时间推移而变化,从而给后来者的使用造成极大的障碍。研究者通过导出数据为纯文本格式,同时以恰当的方式记录缺失值、权重和其他具体的数据编码,可以确保即便统计软件的专门格式发生了改变,数据仍然可用。

——Micah Altman

(赵锋译校)

参考文献

Altman, M., Gill, J., & McDonald, M. P. (2003). *Numerical methods in statistical computing for the social sciences*. New York: Wiley.

Gentle, J. (2003). *Computational statistics*. New York: Springer-Verlag.

McCullough, B. D., & Vinod, H. D. (1997). The numerical reliability of econometric software. *Journal of Economic Literature*, *37*(2), 633-665.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer-Verlag.

统计检力(Statistical Power)

用样本数据对总体【Population】值的假设【Hypothesis】做检验时,其中一个可能是:由数据推论维持虚无假设【Null Hypothesis】的判断,然而这个判断却是伪的。这是第二类错误【Type Ⅱ Error】(β)。然而,为了强调检验的正效力而非负效力,特将术语检力定义为第二类错误的反置:统计检验的检力是否定错误假设的概率或$P=1-\beta$。统计检力是极其重要的概念,因为多数社会科学研究是设计来证明某一具体的变量对某种行为有实际的影响。在实际研究中,我们着手找出证据,证明"无效应假设(hypothesis of no effect)"(即虚无假设)确实是伪假设。我们想证明的是:某种具体的治疗是有效的,一种教育方案是可行的,世界事件会影响政治舆论,以及诸如此类。为此,我们就需要有高水平的统计检力。

尽管,当α的值,即犯第一类错误【Type Ⅰ Error】(否定真实假设)的概率,被设定得较低时,统计检力会降低,然而一些好的实验设计【Experimental Design】原则会提高检力:使用大样本,设计可靠的响应测量,选取同质的回答总体,就同一主题运用重复的测量,以及使用标准的、定义清晰的作业内容和作业说明。然而,有一个因素我们无法控制,即统计检验所欲估计的总体参数的真值。虚无假设同真值间的差越大,统计检力就越高。逻辑上(也很幸运的是),我们更

易于判定强(有效)处理的效应而非弱处理的。有些研究者给出了检力函数或检力曲线,以说明统计检验的检力随虚无假设的错误度变化的关系(Cohen, 1988)。相反,一些研究者会求助于早前的研究工作或通过他们自己的前测工作,预先估计研究的试验效应的级别,再确定样本量,以此确保他们的研究能满足预定的统计检力,比如0.90[读者可参看科恩(Cohen,1988)对此的具体说明]。

海斯(Hays,1994)将统计检力比作显微镜的功效。正如一台高倍显微镜有助于增加我们发现某些低倍显微镜无法看到的事物的机会,高检力的检验则有助于否定错误的虚无假设。然而,即使一台高倍显微镜也会受限于所欲发现的对象的级别。同理,在研究中我们欲发现的效应越大,发现此效应的检验的检力也越大。好的研究技术同样可以增加发现效应的可能性。

——Irwin P. Levin

(赵锋译校)

参考文献

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.

Hays, W. L. (1994). *Statistics* (5th ed.). Fort Worth, TX: Harcourt Brace.

统计显著(Statistical Significance)

当某估计得到统计显著时,估计值就可能是非零的,而该值也值得予以特别关注。统计检验有时会用于单变项【Univariate】统计量,如均值;或双变项【Bivariate】统计量,如皮尔逊相关系数【Pearson's Correlation Coefficient】;或多变项统计量,如多元回归【Multiple Regression】方程的斜率系数。例如,若变量X和变量Y间的皮尔逊相关系数在0.05的水平(最常用的显著水平)上统计显著,那么就否定关于研究总体中X和Y无关联的虚无假设【Null Hypothesis】。然而,若该相关在0.05的水平上不显著,我们倾向于认为,至少初步地判断,X和Y的联系不值得进一步研究。因此,统计显著起到门卫的功能,若"通过此门",对此系数就要有进一步的考量;若没能"通过此门",对此系数就无须或暂缓考量。

——Michael S. Lewis-Beck

(赵锋译校)

* 也可参见显著性检验【Significance Testing】。

茎叶图(Stem-and-Leaf Display)

茎叶图由约翰·图基发明,是一种图示化的展示数据分布【Distribution】的方式。茎表示归组值的主要单位,叶表示此单位上值的细分。在图上,需将每个可能的茎值都标记出来,即便其上并无个案。然后,可据此图观察数据分布的聚合和缺口,以及分布是否是偏态的【Skewed】。茎叶图也可以标识出异常的值。要以最恰当的方式呈现数据的分布,就需为茎值选取合适的单位,既不能选得太大,以至于过度压缩数据,也不能选得太小,过度分割茎的单位,从而表现不出数据的分布模式。可将茎值除以2,5

或 10，从而延展所给分布的汇总；也可将茎值乘以 2，5 或 10，从而压缩所给分布的汇总。例如，表 1 给出了一组来自 1990 年代后期英国调查的 30 户家庭的收入值。

若设定茎值的单位为 10，随后由原始值得到的茎和叶的值显示在表 1 的第二列，其茎叶图如图 1 所示。

表 1　某个英国调查的收入分布中的 30 个样本个案及标在茎叶图上的值

收入/英镑	茎值=10 上的茎叶值	茎值=20 上的茎叶值	茎值=50 上的茎叶值
61	61	0s6	0.6
63	63	0s6	0.6
74	74	0s7	0.7
74	74	0s7	0.7
80	80	0.8	0.8
84	84	0.8	0.8
85	85	0.8	0.8
86	86	0.8	0.8
88	88	0.8	0.8
95	95	0.9	0.9
99	99	0.9	0.9
107	107	1*0	1*0
120	120	1t2	1*2
121	121	1t2	1*2
122	122	1t2	1*2
123	123	1t2	1*2
131	131	1t3	1*3
133	133	1t3	1*3
133	133	1t3	1*3
135	135	1t3	1*3
136	136	1t3	1*3
139	139	1t3	1*3
142	142	1f4	1*4
144	144	1f4	1*4
150	150	1f5	1.5
156	156	1f5	1.5
171	171	1s7	1.7
188	188	1.8	1.8
200	200	2*0	2*0
274	274	2s7	2.7

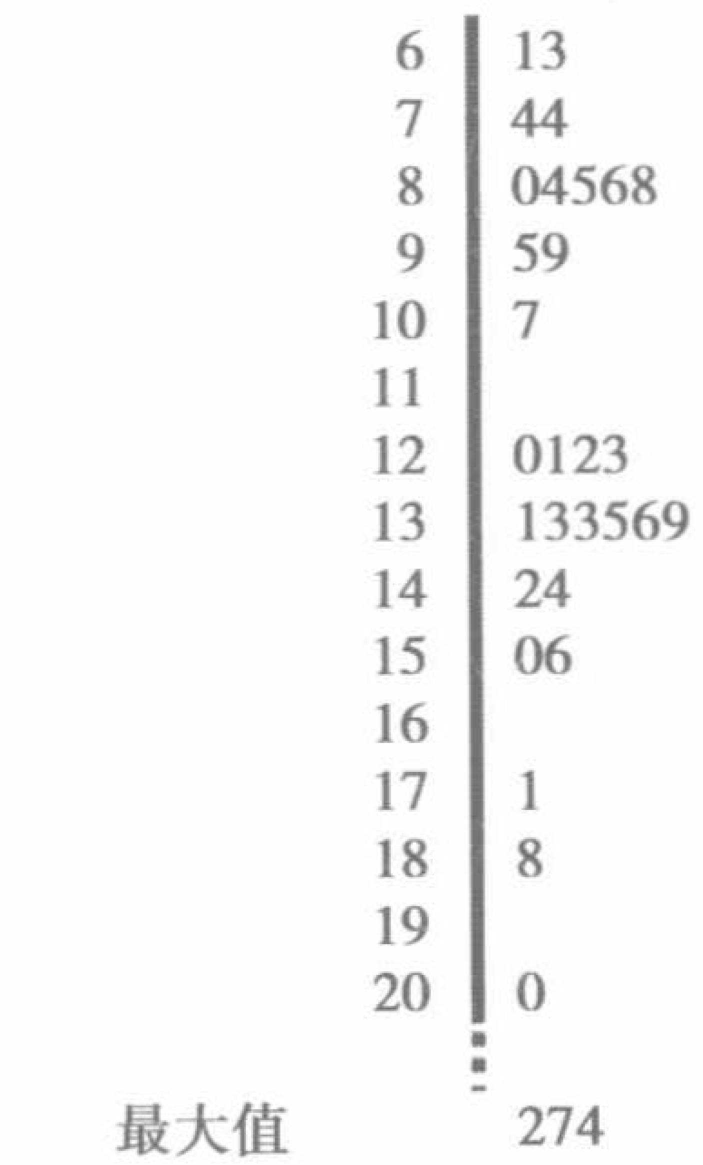


图 1　表 1 中数据的茎叶图

注：茎值单位为 10 英镑；叶值单位为 1 英镑。

由图 1 可知，分布的峰值为 130～139，尽管在值 80 上也有一个较矮的峰。可以看出在值 110，160 和 190 上都有缺口。这可能表示分离的分布上的断点，当然也可能是由于样本量过少造成的。同时，没有值大于 200，小于 274，这并不表示一串空的茎点，反而把茎图变得更长、更散乱，同时，最大值 274 直接被作为异常值【Outlier】标注在图的底端。

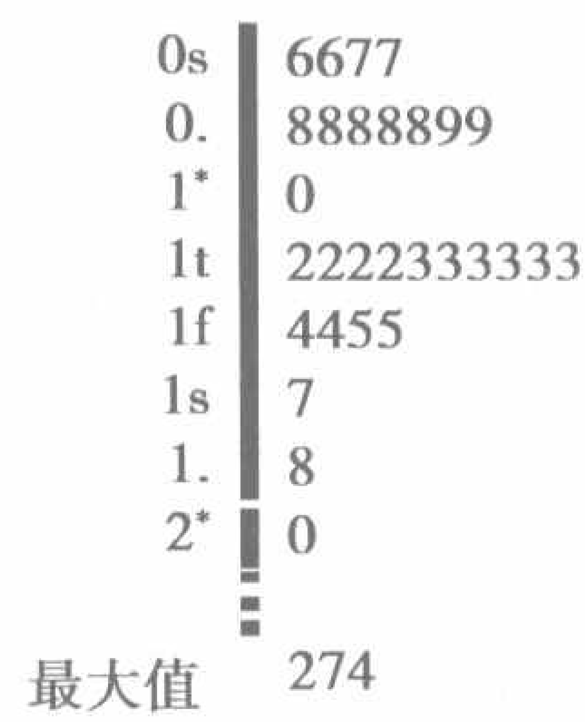


图 2　表 1 中数据的茎叶图

注:茎值单位为 20 英镑;叶值单位涵盖(x_0~x_9)英镑。

在此茎叶图中,原始数据的确切值得以保留:我们知道最小值是 61,最大值为 188。如果不是以 10 等分来分割茎,而是减少茎上分割的数目,原来的数据值就会被简化。例如,若茎上的每个点表示间距 20,那么叶上就只能根据点是否落于 2 个间距为 10 的范围予以区分。图 2 给出表 1 中第三列数值的茎叶图。此处,茎点的数值是一个百位数,其后的符号表示 20 等分所参照的数值范围:"＊"表示数值为 0~19;"t"表示 20~39;"f"表示 40~59;"s"表示 60~79;"."表示 80~99。叶上的数字仅表示两个 10 间距中的那一个包含了该数据。

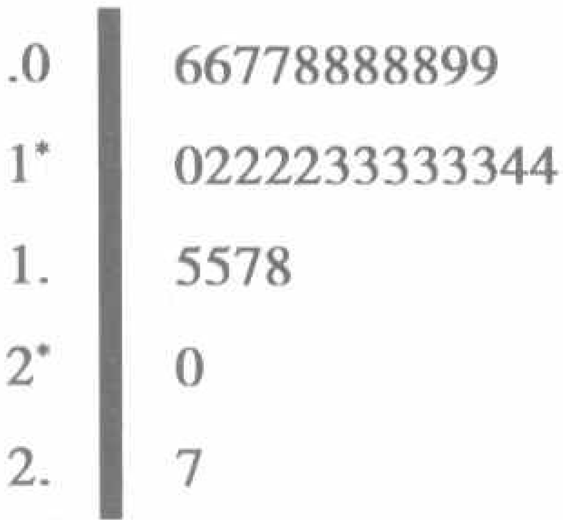


图 3　表 1 中数据的茎叶图

注:茎值单位为 50 英镑;叶值单位涵盖(x_0~x_9)英镑。

这里,尽管分布的形状没有给出更多的细节,但是清楚地显示了两个峰和一个长尾。在此例中,若进一步收缩标尺,如以 50 作为分位,如图 3 所示,茎叶图就无法给出有效的信息了。

——Lucinda Platt

(赵锋译校)

参考文献

Marsh, C. (1988). *Exploring data: An introduction to data analysis for social scientists*. Cambridge, UK: Polity.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

阶梯函数(Step Function)

参见干预分析【Intervention Analysis】。

逐步回归(Stepwise Regression)

给定一组可包含于线性回归【Linear Regression】方程中的可能的解释变量【Explanatory Variable】,当缺乏选择变量先后次序的理论和主观知识时,于是发展出选取合适的变量进入回归方程的自动方法。逐步回归是一组建构模型的算法。这组算法【Algorithm】基于自变量对因变量的回归拟合的贡献大小,或者包含某些变量进入方程,或者将某些变量从方程中剔除。尽管数据驱动的自动模型【Model】选择程序在大部分统计软件都可选用,似乎也很有吸引力,但是由逐步回归选出的模型一般而言缺乏最优特性,并且不能给出正确的统计推论。通常,由逐步算法选取的模型既不是正确的

模型,也不是最佳的拟合模型和基于样本的好的预测模型。

虽然存在多种逐步算法(stepwise algorithms),但是每种算法的逻辑是相同的。逐步算法基于变量的贡献能否满足特定的拟合准则(fit criterion),而逐一筛选某个变量或某组变量是否纳入回归模型。就线性累加模型而言,最常用的准则是保持每一步上残差平方和最小。基于对数似然统计量(log-likelihood statistics)的更一般的方法应用了相同的逻辑,适用于类别【Categorical】数据及其他类型的数据。当预先决定的最大数量的变量选择完成或模型拟合指标的变化小于预先确定的加入或排除额外变量的阈限时,选取“更佳”模型的工作停止。

两种最常用的实现方式是前向选取(forward selection)和后向选取(backward selection)。前向选取更多用在潜在回归项(regressor)的数量(k)接近或超过数据集中观测值的数量(n)。前向选取中,初始模型(initial model)通常只以均值拟合,即

$$Y_i = \beta_0 + e_i$$

在随后的每第 p 步上,从剩余的潜在的解释回归项($x_1, x_2, \cdots, x_k$)中选取变量 $x_{(p)}$ 加入模型,即

$$y_i = \beta_0 + \beta_1 x_{(1)i} + e_i$$

$$\cdots$$

$$y_i = \beta_0 + \beta_1 x_{(1)i} + \beta_2 x_{(2)i} + \cdots + \beta_p x_{(p)i} + e_i$$

使用最小二乘法【Least Squares】估计参数 $\hat{\beta}_J$,p 个变量的残差平方和【Residual Sum of Squares】为

$$\text{RSS}_p = \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \sum_{j=1}^{p} \hat{\beta}_j x_{(j)1} \right)^2$$

在第 p 步,剩余的 $k-p-1$ 个被排除的变量逐一地加入当前模型中,同时,计算和保留这 $k-p-1$ 个回归方程的残差平方和。纳入回归方程中的变量是引起残差平方和的最大增量变化的那个。

$$F = \frac{\text{RSS}_{p-1} - \text{RSS}_p}{\dfrac{\text{RSS}_p}{n-p-2}}$$

对逐步回归的描述可以另表述为,选取程序使用下列等效准则之一,选定进入方程的变量。这些等效准则包括:(a)增加 R^2 最多;(b)得到最大的 t 统计量;(c)与因变量有最大的偏相关(partial correlation)。变量被逐步加入回归方程中,直到变量的增加无法得出显著的 F 统计量,或者达到了使用者选定变量数量的上限(p)。注意,在逐步搜寻的条件下,F 统计量实际上并非得自某个 F 分布【F Distribution】。

后向选取正同前向选取相反,先从包括全部变量的回归方程开始,然后等效地排除变量,(a)减少 R^2 最少的;(b)有最小的 F 统计量或 t 统计量;(c)与因变量有最小的偏相关。某个变量一旦从方程中去除,即不再考虑将其纳入模型中。注意,如果观测值的数量少于变量的数量,要把所有的变量加入回归方程是不可能的。当 $n<k$ 时,或者避免使用后向选取方法,或者研究者在开始计算前预先确定排除至少 $k-n-1$ 个变量。

无论基于理论的还是实践的原因,逐步回归作为模型选择的工具都难以胜任。当同一数据用于模型建立和模型选择以及假设检验【Hypothesis Testing】时,对照由逐步回归选出的最终模型,该方法使得多数标准的统计推论理论无效(Miller,1990)。实际上,由于程序的选择性偏倚【Selection Bias】,选取变量的参数估计会偏大。基于选取模型的统计推论也是不正确的,一方面

因为回归系数的信度区间【Confidence Intervals】会太小;另一方面参数子集的 F 检验会虚假显著,因为残差方差会被低估。而且,逐步回归不仅通常无法找出生成数据的正确模型,甚至不能发现最佳的拟合模型。在极端条件下,如所有潜在回归项一一正交,逐步回归的问题才降至最低;随着变量间的相关【Correlation】增加,相应的问题变得严重。

存在着其他的像逐步回归那样从大量竞争的解释变量中筛选变量的方法。如果多个解释变量确认是少量共同潜在测量的表征(manifestations),那么较佳的做法是把这些潜在测量的综括量(summaries)作为解释变量纳入回归方程。例如,可以通过只用多个解释变量的主成分【Principal Components】作为回归项【Regressors】,以减少它们的维度。如果存在多个解释变量,每组代表竞争的理论,那么最好形成多个非嵌套(non-nested)的假设检验,直接估计竞争理论间差异的显著水平。研究者最好根据理论和数据的实际知识来指导建模,而不要依赖如逐步回归那类通常容易产生误导的自动筛选程序。

——Jonathan Wand

(赵锋译校)

参考文献

Miller, A. J. (1990). *Subset selection in regression*. New York: Chapman Hall.

随机的(Stochastic)

不能完全控制或决定其值的变量是随机的。若某个变量测量一项实验的 N 次试验的结果,同时,在观测前不能决定或预测这些结果,该变量即是随机变量。试验结果的随机变化(random variation)源于机会,此即,随机变量由概率分布定义,而非确切的或决定函数。一般而言,一个随机的因变量 $y=g(\mathbf{X})+\varepsilon$,由决定函数 $g(\mathbf{X})$,简单情况下也可是某个常数和一个随机扰动项或误差【Error】项 ε 构成。

重要性

在社会科学方法论中,随机是一个基本概念,因为所有的因果过程(因变量)都假定是随机的。在自然科学中,因果关系是确定的而非随机的,所以研究者可以通过设计一个控制实验——实施试验,以涵盖解释变量取得的所有值,从而计算方程的参数【Parameter】,确定变量间的关系。当因果关系确定时,实验的试验结果就是决定函数上的值,而不是得自以该函数为中心的概率分布【Probability Distribution】上的一个样本,因为若因变量是非随机的,则不存在抽样误差【Sampling Error】。相反地,社会科学的研究者必须使用准实验设计【Quasi-Experiment】,面对更多的测量【Measurement】限制,考察涉及人类行为的更复杂的因果过程,而不是物理要素的相互影响。因此,社会科学中所有的因果过程被假设为本质上是随机的,即必须通过抽样【Sampling】和统计推论【Statistical Inference】才能考察。就某个随机因变量 $y=g(\mathbf{X})+\varepsilon$,思考预测其值的假设的含义。即便决定部分 $g(\mathbf{X})$ 已知,但是因为随机部分 ε 影响因变量的值围绕给定 $\mathbf{X}$ 下 y 的条件均值[由 $g(\mathbf{X})$ 确定]发生变化,所以预测始终伴随着误差。

假设所有因变量是随机的,一般有三种合理的解释:第一,随机扰动解释了被忽略的无数小的、独特的影响因素。这些被忽略

的影响因素对因变量的净效应(net effect)被假设随观测随机变化。第二,在社会科学中,因变量通常有测量误差,即便不总是如此。第三,人类行为本身是随机【Random】的。在经济、政治和社会条件下,这些影响因素从不完全信息(incomplete information)到复杂决策的策略行为阻止人们以完全确定的方式行事。此项合理性常常基于这样的论证,即社会科学的总体是无限的,因为若人类行为是随机的,即便当下的历史也只是所有可能发生的历史的一个样本。

在重复的样本中,解释变量常常假定是非随机的或固定的。当解释变量是随机的时,关键问题是它们是否同干扰项相关。如果相关,它们会导致有偏的参数估计【Parameter Estimation】;如果无关,解释变量代表辅助的回归项【Regressors】(即回归项的边缘概率分布不依赖感兴趣的参数)。在动力模型(dynamic models)的情况下,这一条件被称为弱的外部性(exogeneity)。当随机回归项是辅助的或弱外部性的,可以通过调整它们的观测值从而将其作为固定项对待。

历史

随机的(stochastic)一词来自希腊词"stokhos",意为"靶心"或目标。掷飞镖过程产生的结果可以模型化为围绕靶心的一个概率分布。随机结果的正式理论始于布莱世·帕斯卡和皮埃尔·费马在 1654 年的关于骰子游戏的一系列信函往来。这组信函提出了概率【Probability】理论的许多基本概念。然而,随机过程理论的发展却始于 20 世纪早期,直到安德烈·马尔科夫提出了马尔科夫链的概念。马尔科夫链【Markov Chain】是一个随机变量序列,在该序列中,未来的变量的值(状态)取决于当前变量的值(状态),但却独立于序列中当前变量的值(状态)所来自的前向变量的值(状态)的影响。将这一概念从离散扩展到一个连续的状态空间(可能的变量值的集)定义了一个一般的随机过程(也称为马尔科夫过程)。

应用举例

思考抛硬币这样的随机事件,其中 H 和 T 表示两种可能的结果。设若硬币是公正的,两个事件是等可能的,即 $P(H)=P(T)=0.5$。不过已知事件结果的概率分布并不能使人们确切地预测任意一次试验的结果。假设 p_N 是 N 次试验中结果为 H 的比率。当 $N=6$ 时,p_N 等于真实参数($\pi=0.5$)的似然仅为 31.25%。事件的随机性质使得 p_N 围绕真实值 π 变化,从而妨碍参数计算,需要使用统计推论。

在回归分析的条件下,请思考另一个例子。假设有一感兴趣的因果过程,被模型成一组外生变量和一个随机扰动项的线性函数,记为 $y=\mathbf{X}\beta+\varepsilon$,其中 $\mathbf{X}$ 是外生变量矩阵,β 是未知参数的向量,ε 是扰动项。函数关系的决定部分记作 $\mathbf{X}\beta$,而 ε 表示随机部分。回归分析使用样本数据估计 β,把因变量 y 的变化分解成可以由 $\mathbf{X}$ 解释的部分(即决定函数引起的变化)和可归之于 ε 而未解释掉的部分(即随机扰动项引起的变化)。作为因变量的函数,β 的估值是一个带有某个抽样分布【Sampling Distribution】的随机变量,给定 $\mathbf{X}$ 对 y 的预测是有误差【Error】的,因为其不确定性既来自随机扰动项,也来自参数的共同估计。为简化统计推论,常常在普通最小二乘估计【Ordinary Least Squares Estimation】估计中假定随机扰动项的概率分布形式。

——Harvey D. Palmer

(赵锋译校)

参考文献

Kennedy, P. (1998). *A guide to econometrics* (4th ed.).Cambridge: MIT Press.

Zaman, A. (1996). *Statistical foundations for econometric techniques*. New York: Academic Press.

分层样本(Stratified Sample)

抽样调查【Survey】的样本通常在设计时需要考虑有关总体【Population】结构的知识。这可能包括,试图确保样本【Sample】在某些重要的方面同总体具有同样的结构,或者,试图有意识地偏离总体结构,从而增加某些子群体的代表性。此类样本设计被称作分层抽样(stratified sampling),而实施该设计的结果即为一个分层样本【Stratified Sample】。

有两类分层抽样:成比例的(proportionate)和不成比例的(disproportionate)。成比例分层抽样包括控制样本在每一层上的比例,使其等于总体的比例。如果层与调查的量度相关,这样的设计就会具有增进调查估计值精度的效应(见设计效应【Design Effects】)。通过创造清楚的层,再分别从每个层中抽样的方法,或者是,先根据某个有意义的顺序,排列抽样框【Sampling Frame】的单位,再从排列的列表中系统地抽样(见系统抽样【Systematic Sampling】)的方法,都可以实现成比例的分层。这两种方法分别被称作“清楚的分层抽样”和“隐含的分层抽样”。成比例的分层不可能对估计值的精度有相反的效应。最不利的情况是所获得的层同某个特定的调查量度完全不相关,这时,分层对精度就没有任何效应,而所得的样本也等价于一个简单随机样本(见简单随机抽样【Simple Random Sampling】)。同时,选取的概率不受成比例分层的影响,所以不会由于拙劣的层的选择而引入偏倚【Bias】。

不成比例的分层需要在不同的层上,使用不同的抽样分数【Sampling Fraction】。其目的时常在于,对那些需要独立估计的一个或若干重要的子群体,增加其样本量。这时,不成比例的分层通常会减少与总的研究总体相关的估计值的精度(见设计效应【Design Effects】),但是会增加被过量抽取的某层或几个层的估计值的精度。然而,当各层间内在的变异度相差很大,且各内在变异度可以被进一步地估计时,不成比例的分层也可以提高总的总体估计值的精度。在这种情况下,可以使用内曼的最优分配(Neyman optimum allocation)规则,以取得最大的精度,

$$\frac{n_h}{N_h} \propto \frac{s_h}{\sqrt{C_h}} \qquad (1)$$

最优分配规则陈述,在层 h 上的抽样分数$\frac{n_h}{N_h}$应当同层的标准差 s_h 成正比,而同层上数据收集的单位成本的平方根 C_h 成反比。

在定量调查中,成比例的或不成比例的抽样选择是研究设计的一个重要且需要详细说明的部分。而在定性研究中,因为目的不在于量化样本同总体间的关系,所以成比例抽样和不成比例抽样的区分无关紧要。但是,如果样本被有意识地设计,以反映总

体结构的某些特征，那么一个样本仍然可以被认为是分层的。

——Peter Lynn

（赵锋译校）

参考文献

Moser, C. A., & Kalton, G. (1971). *Survey methods in social investigation* (2nd ed.). Aldershot, UK: Gower.

Scheaffer, R. L., Mendenhall, W., & Ott, L. (1990). *Elementary survey sampling*. Boston: PWS-Kent.

Stuart, A. (1984). *The ideas of sampling*. London: Griffin.

关联强度（Strength of Association）

关联强度显示两个变量共变【Covary】的程度，以及自变量【Independent Variable】影响因变量【Dependent Variable】的程度。

关联强度将两个变量【Variable】的关系概括成一个总和的统计量【Statistic】，其取值为0~1。关联度的绝对值越大，两个变量间的联系越强。据所使用的关联模型，取值1表示变量间完全关联，取值0表示变量间无关系。就定序和定距数据而言，当一个变量较大的值对应另一变量较小的值时，则有负值出现。

两个变量间的关联强度可以对应于实验【Experiment】中的处理【Treatment】差异。在此情形下，可以使用非对称量度【Asymmetric Measures】予以测量。尽管通常会用对称的系数测量，不过对称的系数不依赖自变量的选择，其取值相同。

许多关于关系的统计量度已被提出，但是它们在估计关联强度上是有区别的。“好的标准”是用于定距数据的皮尔逊的 R 统计量，其值的平方和表示一个变量的方差可以被另一个变量通过线性【Linear】预测而予以解释的比例。还有一些用于非定距变量的量度也可以相似地解释为误差减少比例【Proportional Reduction of Error】（Costner, 1965）。其中一些量度当变量间存在着强的单调【Monotonic】关系时，会取值1，而另一些[如耶尔的 Q]仅要求弱的单调关系就可取值1，即一个变量上的较大值对应另一变量上相同或较大的值时。

当变量间没有关系时，皮尔逊的 R 统计量，如同其他此类系数，例如戈德曼和克鲁斯卡尔的 τ 和耶尔的 Q，可以解释成两变量间统计独立【Independence】。韦斯伯格（Weisberg, 1974）就一些系数区分了其他的零条件。例如，定距变量的 Lamda（λ），当自变量上的每个类别都对应于因变量的众类时，有零值。

尽管一般认为系数大于0.70就表明变量间的关联相当大，而小于0.30则表明关联相当小，然而没有严格的标准用以估计一个关联的大小。

关联强度的概念可以被推广至多于两个变量的情形。例如，典型相关【Canonical Correlation】表示两组变量间的关联强度。统计量 R 方【R-Squared】则表示由回归方程【Regression Equation】所刻画的关联强度。

——Herbert F. Weisberg

（赵锋译校）

参考文献

Costner, H. L. (1965). Criteria for measures of association. *American Sociological Review*, *30*,

341-353.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, *49*, 732-764.

Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association*, *53*, 814-861.

Weisberg, H. F. (1974). Models of statistical relationship. *American Political Science Review*, *68*, 1638-1655.

结构系数(Structural Coefficient)

术语结构系数常用于路径分析【Path Analysis】和结构方程模型【Structural Equation Modeling】(Duncan, 1975)的讨论,以示其区别于回归分析的讨论术语——回归系数。为便于讨论这一术语,一个常用的做法是以路径图的形式来呈现结构方程组。图 1 给出用于学生水平的科学成绩模型的路径图。这幅图曾在卡普兰(Kaplan,2000)的书中用于讨论路径分析。路径图是特别有效的绘图工具,因为,如果绘制精确,那么图和结构系数组间就存在一对一的关系。

图 1 中呈现的模型结构方程系统可以写成

$$\mathbf{y} = \boldsymbol{\alpha} + \mathbf{B}\mathbf{y} + \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\varsigma} \qquad (1)$$

其中,**y** 是观测到的内生变量【Endogenous Variable】的向量,**x** 是观测到的外生变量【Exogenous Variable】的向量,**α** 是结构截距的向量,**B** 是联系着每个内生变量的结构系数矩阵,**Γ** 是将内生变量联系到外生变量的结构系数矩阵,而 **ς** 是干扰项的向量,其中 Var(**ς**)= **Ψ** 是干扰项的协方差矩阵。最后,设 Var(**x**) = **Φ** 是外生变量间的协方差矩阵。

B 和 **Γ** 中的元素是结构系数,代表变量间的关系。这些矩阵中零和非零元素的模式则由潜在的实质理论设定。例如,如图 1 的模型所示,**B** 中的一个结构系数可能是联系 10 年级的科学成绩(SCIGRA10)同学生认为他们的科学课所具有的挑战性(CHALLG)间的路径。**Γ** 中的一个结构系数可能是联系 SCIGRA10 与社会经济地位(Socioeconomic Status,SES)间的路径。

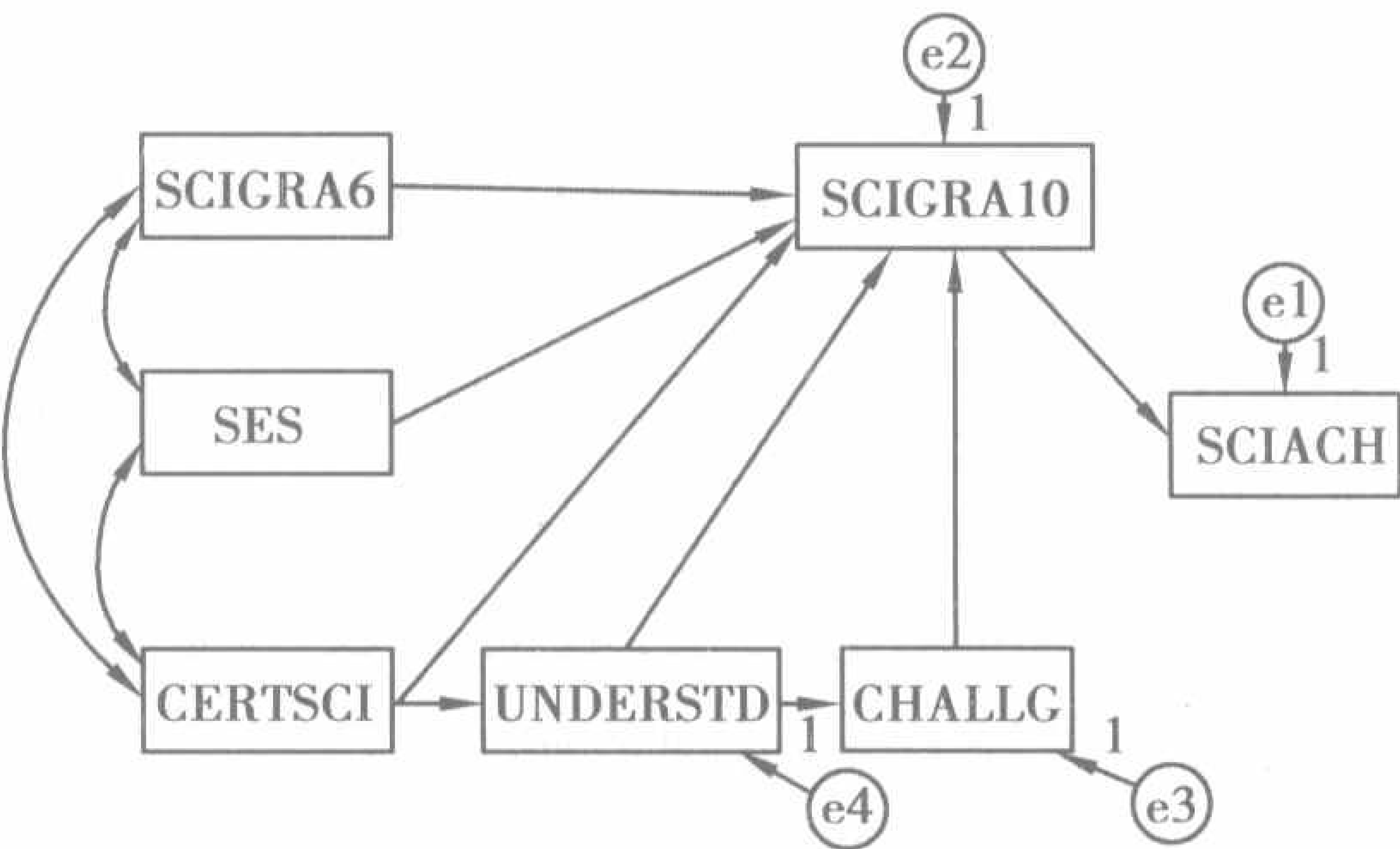


图 1 教育指标模型:初始设定
资料来源:Kaplan(2000)。

在 **B** 和 **Γ** 中,我们可以分辨出三类结构系数。第一类是那些待估的系数。这类系数常被称作自由参数。因此,在图 1 的模型中,可见的路径就是自由参数。

第二类参数是在估计中事先被设定为常量的参数。这些参数常被称为固定参数。大多数情形下,固定参数被设定为等于零,以表示两个变量间不存在关系。根据图 1 的模型,我们在理论上设定科学成绩(SCIACH)同 SES 间没有直接关系。因此,这一结构系数就固定为零。最后,在模型中,也可限定某些参数等于其他参数。这类元素被称为限定参数。限定参数的一个例子可能是要求 SCIGRA6 对 SCIGRA10 的效应等于 SES 对 SCIGRA10 的效应。

结构系数还有三项另外的性质值得注意。第一,如同一般的回归系数一样,结构系数也可以做显著性检验。第二,若结构系数是标准化的,它们就被称为路径系数(Duncan, 1975; Wright, 1934)。第三,结构系数可以相乘和相加,从而形成间接效应和总效应的系数(Duncan, 1975)。

——David W. Kaplan

(赵锋译校)

参考文献

Duncan, O. D. (1975). *Introduction to structural equation models*. New York: Academic Press.

Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Thousand Oaks, CA: Sage.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, 5, 161-215.

结构方程模型(Structural Equation Modeling)

结构方程模型(SEM)可以定义为一组方法论,其目的在于根据一组较少数量的、由假定的潜在概念或理论模型所确定的"结构"参数来表示关于观测数据的均值、方差和协方差的假设。从历史的角度看,结构方程模型源自两个独立的统计传统的结合。第一个传统是自计量心理学发展而来的因子分析。第二个传统是主要从计量经济学发展而来的联立方程模型。联立方程模型在遗传学领域也有其早期的历史,并以路径分析【Path Analysis】的名义被引入社会学领域。关于结构方程模型的历史更加翔实的论述可以参见卡普兰(Kaplan, 2000)的文章。

一般模型

由约雷斯括格(Jöreskog, 1973)所概述的一般结构方程模型包括两个部分:一部分是经由联立方程系统连接了各个潜变量的结构部分;另一部分是经由一个限定的(证实性)因子模型,将观测变量同潜变量连接在一起的测量部分。模型的结构部分可以写作

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\varsigma} \qquad (1)$$

其中,$\boldsymbol{\eta}$ 是内生(准则)潜变量的向量,$\boldsymbol{\alpha}$ 是结构截距的向量,$\boldsymbol{\xi}$ 是外生(预测)潜变量的向量,$\mathbf{B}$ 是联系每一个内生潜变量的回归系数矩阵,$\boldsymbol{\Gamma}$ 是将内生变量和外生变量联系在一起的回归系数的矩阵,而 $\boldsymbol{\varsigma}$ 是干扰项向量。

无论对内生变量还是外生变量,潜变量都可由测量方程与观测变量相连接。这些方程可以定义为

$$\mathbf{y} = \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\varepsilon} \qquad (2)$$

和

$$\mathbf{x} = \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta} \qquad (3)$$

其中,$\boldsymbol{\Lambda}_y$ 和 $\boldsymbol{\Lambda}_x$ 是各个方程的"因子负荷(factor loadings)"阵,而 $\boldsymbol{\varepsilon}$ 和 $\boldsymbol{\delta}$ 是各方程的特性(uniqueness)向量。此外,一般模型还设定了 $\boldsymbol{\xi}$,$\boldsymbol{\varsigma}$,$\boldsymbol{\varepsilon}$ 和 $\boldsymbol{\delta}$ 的方差和协方差,分别记作 $\boldsymbol{\Phi}$,$\boldsymbol{\Psi}$,$\boldsymbol{\Theta}_\varepsilon$ 和 $\boldsymbol{\Theta}_\delta$。一个原型的结构方程模型的路径图如图 1 所示。

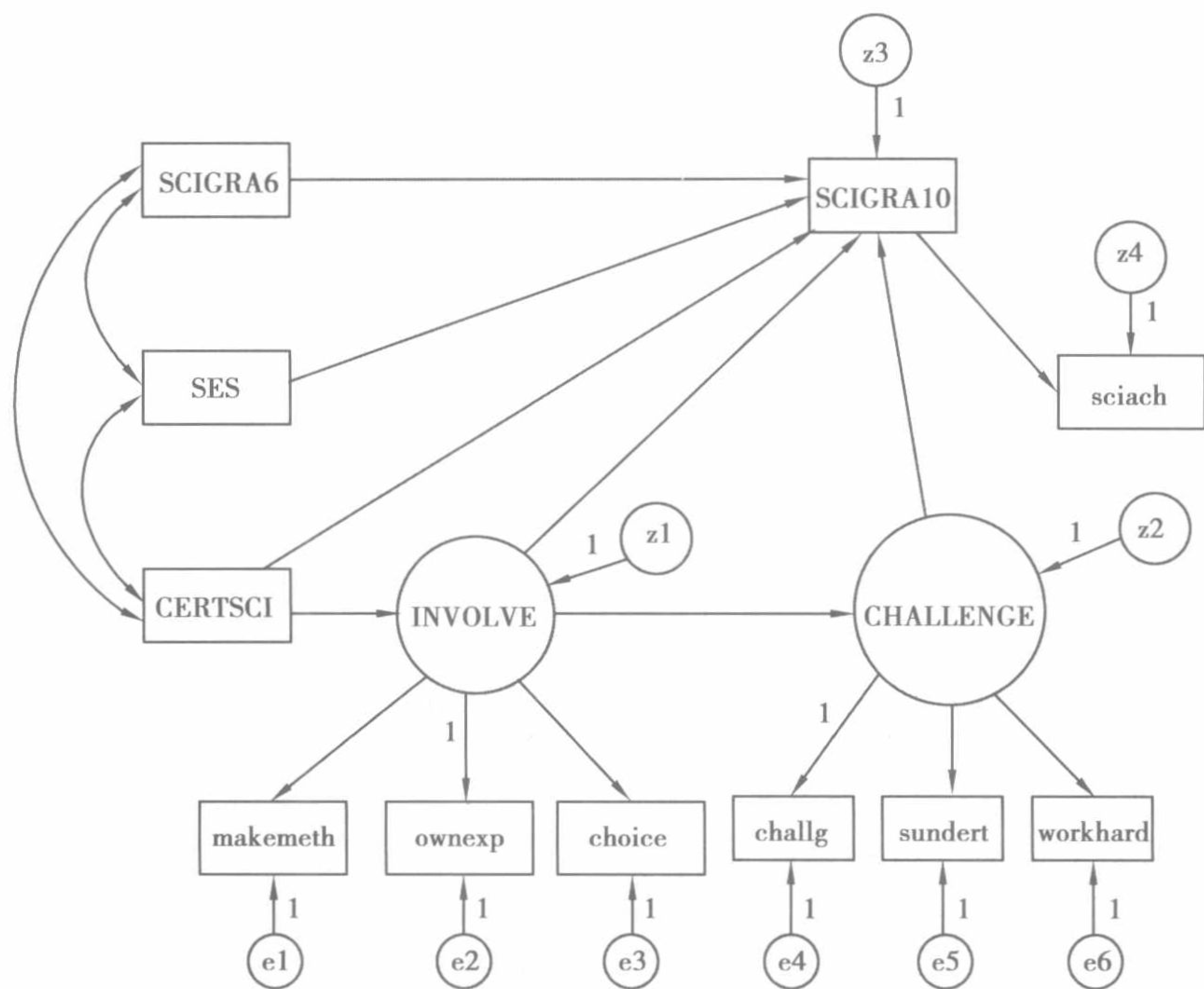


图 1 科学成绩模型的结构方程模型

资料来源:Kaplan(2000)。

结构方程模型作为一种方法论,其设计的首要目的在于检验实质的理论。因此,理论必须得到充分的发展,以便能够指明某些构念(constructs)不影响另一些构念,某些变量并没有负荷于某些因子上,以及某些干扰项和测量误差不存在协变。这就隐含着 $\mathbf{B}$, $\mathbf{\Gamma}$, $\mathbf{\Lambda}_y$, $\mathbf{\Lambda}_x$, $\mathbf{\Phi}$, $\mathbf{\Psi}$, $\mathbf{\Theta}_\varepsilon$ 和 $\mathbf{\Theta}_\delta$ 中的某些元素可以根据假设固定为零。剩余的参数即为自由的待估参数。从这个角度讲,结构方程模型可以被视作更一般的"协方差结构模型(covariance structure model)"的一个特例,而协方差结构模型可以定义作 $\sum = \sum(\mathbf{\Omega})$,其中 $\sum$ 是总体的协方差阵,$\sum(\mathbf{\Omega})$ 则是包含了所有模型参数的"参数向量(parameter vector)" $\mathbf{\Omega}$ 的待估的矩阵函数。

结构方程模型的判定

欲估计模型的参数,一项先决条件是确定参数是否是可判定的(identified)。判定性(identification)的定义可通过自协方差结构模型的视角思考问题而给出。总体协方差阵 $\sum$ 包含了总体的方差和协方差,同时,假定这些方差和协方差服从由 $\mathbf{\Omega}$ 中一组参数所定义的模型。我们知道,$\sum$ 中的方差和协方差可根据样本协方差阵 $\mathbf{S}$ 中的它们的样本对应部分予以估计,而直接使用公式就可以计算样本的方差和协方差来得到样本的协方差阵 $\mathbf{S}$。因此,$\sum$ 中的参数是可判定的。先于估计所要确定的是,$\mathbf{\Omega}$ 中的未知参数是否可以根据 $\sum$ 中的元素加以判定。

我们说 $\mathbf{\Omega}$ 中的元素是可判定的,如果

它们可根据协方差阵 Σ 的元素唯一地表示出。如果 $\mathbf{\Omega}$ 中的所有元素是可判定的，则模型是可判定的。用以确立模型判定性的多个必要和充分条件在卡普兰(Kaplan, 2000)的书中有具体的描述，其中关于判定性的秩条件和序条件来自计量经济学传统。

结构方程模型的估计

在结构方程模型中，参数估计的问题考量的是在满足模型的固定参数所设定的限定条件下，获得模型的自由参数的估计值。其目标在于，将模型应用于样本协方差阵 $\mathbf{S}$，而它是被设定服从某个结构方程模型的总体协方差阵 Σ 的估计值。当估得的协方差阵 $\hat{\Sigma} = \Sigma(\hat{\mathbf{\Omega}})$ 同样本协方差阵 $\mathbf{S}$ 的差值尽可能小时，可得到参数的估计值。

参数的估计【Estimation】需要在多种“拟合函数(fitting functions)”间选取适当的拟合函数。拟合函数具有下列属性：(a)函数的值大于或等于 0；(b)当且仅当模型完美地拟合数据时，函数值等于 0；(c)函数是连续的。在满足观测数据的多变量正态性条件下，可以使用最大似然估计【Maximum Likelihood Estimation】来进行参数估计，若放宽关于数据分布的假定，那么一般加权最小二乘估计可用于连续的非正态的和/或定类的数据。

检验

一旦模型的参数得到估计，就可以检验模型对数据的全拟合(global fit)。虚无假设陈述，设定的模型对总体成立。对应的替代假设是：总体协方差阵是任意的对称(及正定义)矩阵。在最大似然估计的条件下，用以检验模型拟合总体的虚无假设统计量被称为似然比【Likelihood Ratio, LR】检验【Test】。似然比检验的大样本抽样分布是卡方分布，其自由度由总体协方差阵中的非冗余元素的数量同模型中自由参数的数量的差值给出。

除了用以检验模型是否确实地在总体中成立的全检验，人们还可以检验模型中个别的固定参数和自由参数。这些检验包括 Lagrange 乘数检验(也称为修正指标)和 Wald 检验。前者验证若设定某个固定参数为自由参数是否会增进模型的拟合，而后者检验若将某个自由参数设定为固定参数是否会降低模型的拟合。这些方法的使用策略在萨里斯等人(Saris, 1987)和卡普兰(Kaplan, 1990)的书中都有详细的讨论。

模型拟合的备选方法

超过 20 年的时间里，人们致力于替代指标的发展，而这些指标则能提供关于结构方程模型拟合相对不同的视角。这些指标的发展，是由于部分地认识到似然比卡方统计量对大样本量的敏感性所受到的鼓动。由此，人们发展了相应的指标，用以将所提出的模型的拟合同变量完全独立的基准模型相比，还包括有增加了一个针对估计太多参数的处罚函数(penalty function)的指标。例如，非标准的拟合指标(non-normed fit index)和比较拟合指标(comparative fit index)。

其他种类指标的发展动机源于需要重新思考对总体的精确拟合的检验观念——在大多数实践的情形中，有些人认为这一想法不切实际。可用于检验对总体的近似拟合的指标是根均方误(root mean square error)的近似。最后，人们还发展了另一类替代指标，其关注点在于模型交叉验证的适当性(cross-validation adequacy)。其中包括 Akaike 信息准则(Akaike Information Criterion)和期望的交叉验证指标(Expected Cross Validation Index)。值得指明的是，对这些指标的应用

并不是不存在着争论(例证可参见 Kaplan, 2000;Sobel & Bohrnstedt,1985)。

SEM 的潜在统计假定

如同所有统计方法,结构方程模型也依赖一组潜在的统计假定。与结构方程模型相关的主要统计假定包括多变量正态性、非系统的缺失数据、充分大的样本量,以及正确的模型设定。

多变量正态性

潜在于结构方程模型的标准应用之下的一项基本假定是观测值抽取自连续的、多变量正态的总体。这一假定对于最大似然估计尤为重要,因为最大似然估计量直接得自针对多变量正态分布的表达式。如果数据服从连续的多变量正态分布,最大似然就可以获得最佳的渐近属性,即估计值是正态的、无偏的和有效的。

关于非正态性对估计值、标准误和模型拟合检验的影响已经有充分的认识。目前的文献认为非正态性不会影响模型参数的估计值。相反地,相对于经验的估计值的标准差,标准误似乎会被低估。就拟合优度而言,研究表明非正态性会导致似然比卡方统计量的实质的高估,而高估的程度似乎同模型自由度的数量相关。

非正态数据的估计量

结构方程模型中的一项主要的突破在于发展出了那些不基于连续多变量正态数据假定的估计方法。布朗(Browne,1984)根据加权最小二乘理论,发展了一个渐近的免分布的(Asymptotic Distribution Free, ADF)估计量,其中加权矩阵采用了一种特定的形式。具体而言,当数据不是多变量正态的,可以扩展加权矩阵以整合关于数据的偏斜和峰态的信息。模拟研究显示,同最大似然【Maximum Likelihood,ML】和广义最小二乘【Generalized Least Squares,GLS】相比,ADF估计有相当好的表现;然而,如果模型相对于样本量而言过大(根据模型的自由度),ADF 则会遇到计算上的困难。在应用实际中,这些约束限制了 ADF 的使用。

与 ADF 估计量相关的另一个困难是社会和行为科学的数据很少是连续的。毋宁,我们会更多地遇到定类的数据,而且常常在具体的分析中发现有多种测量层次的混合。当遇到这样的情形时,可能同时会背离连续性和多变量正态性的假定。通过提出用以估计含有测量层次混合的模型的统一方法,木滕(Muthén,1984)为估计此类现实条件下的结构方程模型做出了主要的贡献。

缺失数据

一般而言,统计方法如结构方程模型,假定每个分析单位都有完整的数据。然而,出于多种原因,用以分析的单位可能在一个变量或多个变量上有缺失值。有许多特定的方法可以处理缺失数据,包括整条删除【Listwise Deletion】和配对删除【Pairwise Deletion】。然而,这些方法都假定,缺失数据【Missing Data】因有缺失数据的变量,或任何模型中的其他变量都不相关,即完全随机的缺失(missing completely at random, MCAR)——最理想情形下的假定。用以处理缺失数据的更成熟的方法来自利特尔和鲁宾(Little & Rubin,1987)的核心思想。

使用利特尔和鲁宾的框架,艾利森(Allison, 1987)和木滕、卡普兰及霍利斯(Muthén, Kaplan, & Hollis,1987)同时作出了突破,给出了结构方程模型背景下处理缺失数据的模型(model-based)方法。他们说明了当缺失数据是随机的缺失(missing at random,MAR)时,比 MCAR 假定更放宽的假定,应当如何使用结构方程模型。

同木滕等人的近似估计量有关的主要限制条件是,它限定于只能对相对少数量的、具有确定模式的缺失数据予以模型。在社会和行为科学应用中,少数量,又具有确定模式的缺失数据是非常罕见的。然而,近年来,已经能够将基于 MAR(Little & Rubin,1987)的模型缺失数据的方法加入标准的 SEM 软件中,同时还能缓解木滕等人的方法带来的问题。具体而言,在假定 MAR 的不完全数据条件下,已经可以将插值法用于结构模型参数的最大似然估计。在 MCAR 和 MAR 条件下,同成列剔除法和配对剔除法相比,模拟研究已经证实了这些方法的效率。目前,在统计软件 LISREL,AMOS 和 Mplus 中,已经可以使用 MAR 条件下的最大似然估计。

设定错误

结构方程模型还假定模型被正确地设定。设定【Specification】错误是指由结构方程模型确定的方程系统中,在任一方程上存在着相关变量的遗漏。这既包括测量模型的方程组,也包括结构模型的方程组。

1980 年代中期,关于结构方程模型中设定错误的问题有大量的研究。根据卡普兰和其他人的基础工作,一般的发现是遗漏变量的设定错误可能导致实质的参数估计偏倚。在抽样变化的条件下,对少量的设定错误而言,标准误的估计是相对稳健的。然而,在模型中自由参数的检验会以这样的方式受到影响,即某个参数的设定错误可能扩散开来,从而影响模型中其他参数的检验效力。

“第二代”结构方程模型的现代发展

多层次结构方程模型

近年来,人们一直致力于将多层次【Multilevel】模型同结构方程模型整合在一起,以便提出一个一般的方法论,既能用于解决测量误差的问题,还能解决多阶段抽样的问题。

多层次结构方程模型假定组内的内生变量和外生变量的水平在不同群组单位间变动。而且,人们可以设定一个模型,既假定它在群组间的水平上成立,同时,用它解释组内变量在群组间的变化。关于多层次结构方程模型的概述和应用可以参见卡普兰(Kaplan,2000)的著作。

结构方程模型视角下的增长曲线模型

多年来,一些研究者一直致力于推荐应用增长曲线模型【Growth Curve Modeling】来研究个体间在变化上的差异。增长模型的设定可以被视作多层次线性模型的子类(Raudenbush & Bryk, 2002),其中水平-1 代表同一个体(intra-individual)在初始位置和增长上的差值,而水平-2 是个体间差值的函数,模拟了不同个体初始位置和增长参数。最近的研究则表明如何利用现有的结构方程模型软件将增长曲线模型整合进结构方程模型框架之中。相对于增长曲线模型,结构方程模型框架的优点是在根据实质的兴趣设定模型时,具有非常大的灵活性。

最后,近年来,木滕(Muthén, 2002)及其同事还推进了增长曲线模型同潜类模型的结合,以便分离出由独特的增长模式所确定的独特总体。这一方法论构成了增长混合模型(growth mixture modeling)。增长混合模型的扩展,称为一般增长混合模型(general growth mixture modeling),容许根据个体的特征来预测潜在的类别归属。

——David W. Kaplan

(赵锋译校)

参考文献

Allison, P. D. (1987). Estimation of linear models

with incomplete data. In C. C. Clogg (Ed.), *Sociological methodology 1987*. San Francisco: Jossey-Bass.

Browne, M. W. (1984). Asymptotic distribution free methods in the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62-83.

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85-112). New York: Academic Press.

Kaplan, D. (1990). Evaluation and modification of covariance structure models: A review and recommendation. *Multivariate Behavioral Research*, *25*, 137-155.

Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Thousand Oaks, CA: Sage.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: John Wiley & Sons.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*, 115-132.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, *51*, 431-462.

Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, *29* (1), 81-117.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Saris, W. E., Satorra, A., & Sörbom, D. (1987). The detection and correction of specification errors in structural equation models. In C. C. Clogg (Ed.), *Sociological methodology 1987* (pp. 105-129). San Francisco: Jossey-Bass.

Sobel, M., & Bohrnstedt, G. W. (1985). Use of null models in evaluating the fit of covariance structure models. In N. B. Tuma (Ed.), *Sociological methodology 1985* (pp. 152-178). San Francisco: Jossey-Bass.

结构语言学(Structural Linguistics)

结构语言学不是把语言作为过程,而是把语言作为结构或系统来研究。1960 年代,它是结构主义【Structuralism】在社会科学和人文学科领域内快速发展的基础,它的影响可以追溯到 1990 年代的许多哲学和理论争论中。

结构语言学的创立者是费尔迪南·德·索绪尔。他与涂尔干的写作几乎同时,而他把语言作为一个社会事实,即存在于一个语言社区的个人成员之中和之上,且强加于成员之上的某种东西。我们不能通过去考察个体言谈行为来研究或理解语言,因为个体言谈行为是独特的,不能作为科学的客体。我们必须审视使得言谈行为成为可能的语言。索绪尔认为一种语言由一定有限数量的发生和将这些发生组合在一起的规则构成,不只是语法的规则,还包括那些在更基础的水平上工作的规则(Saussure, 1974)。语言的基本单位是符号,而符号由一个能指(即发声或纸上的记号)和一个所指(即概念)构成。需要强调的是,所指是概念而不是客体。词和物之间的关系是约定俗成的,而不是必然的;也就是,如果我们想要成为我们的语言社区的一部分,我们就不得不接受此类关系,例如"腿"这个词和我们身体下部的(最)长的东西之间的关

系。实际上,它们可以被用于称呼任何事物。这种观念既打破了我们在词和物之间的常识关系,也为 20 世纪社会理论和哲学的“语言学转向”做出了贡献。

语言符号与其他符号一起被组合进一个系统中,因此,我们可以通过一种伴随(concomitant)变换,或改变每一个要素,直至我们发现一个能够改变符号意义的要素,来发现一种语言的意义要素。例如,如果我们把“cat”中的“c”改成“m”,我们得到一个非常不同的意义,从而把对立“c/m”确立为一个意义要素。这些要素可以在两条轴线上分析:组合(syntagmatic)层面,即考察一个句子中何者可以随在何者之后的规则;词形变化(paradigmatic)层面,即考虑在一个句子中,何者可以被何者替代。

这意味着在传统的语言系统中,每个符号在与其他符号的关系中获得自己的意义;意义不存在于词中,而存在于词之间。简言之,结构主义运动的理念是我们经验的世界是潜在结构的产物。这一理念在分析文化系统时特别有用,而它的变体则从克劳德·列维-斯特劳斯(Lévi-Strauss, 1969)的皇皇巨著《亲属制度的基本结构》(*The Elementary Structures of Kinship*)直到威廉·赖特(Wright, 1975)对不同类型的西方电影的潜在叙事结构的分析。

——Ian Craib

(王玥译　赵锋校)

参考文献

Lévi-Strauss, C. (1969). *The elementary structures of kinship*. London: Eyre and Spottiswoode.

Saussure, F. de (1974). *Course in general linguistics*. London: Fontana/Collins.

Wright, W. (1975). *Six guns and society*. Berkeley: University of California Press.

结构主义(Structuralism)

结构主义根源于这一理念,认为语言是从语言系统中并置的要素来获取意义的。作为社会科学中向文本转型的先行者,通过研究系统的规则和潜在于文本的模型,结构主义为文化研究提供了严密客观的特征。结构主义分析的标志在于:强调语言是使得个人话语和行动变得有序的社会权力。尽管结构主义被批评为反人性的和决定论的,其核心观点仍旧被包含到符号学【Semiotics】、后结构主义【Poststructuralism】、后现代主义【Postmodernism】和结构理论之中。

结构语言学【Structural Liguistics】之父索绪尔提出,研究语言的正确方法是去研究语言系统潜在的规则和模型,因为潜在的规则和模型使得真实生活的话语运转成为可能。索绪尔忽视语言的历史演变和个人使用的重要性,认为语言是集体规则(语言系统)的静止系统,而不是一系列的话语行动(言语)。而且,语言的意义要在语言系统的要素之间的关系中发现,而不是任何语言之外的考量。这些智识进展为以集体的、系统的、规则为本的语言学研究和其他的文化研究铺平了道路。

作为文化系统自主性的强纲领,结构主义被认为是对唯物论视角的唯心论【Idealism】替代,如马克思主义。随后人类学家将这一方法扩大范围到神话、宗教、亲属系统。列维-斯特劳斯发现神话可以通过探查结构化叙事形式和社会生活本身的潜在次序来研究。最有意义的方法论概念是二元对立,在一组词汇中一种不断再现的两极化差别——如男—女、生—熟、神圣—世

俗、纯洁—危害以及自己—他人——列维-斯特劳斯将其归因于思维本身的结构。这些对立关系通过差别和排斥原则,为经验性的基础资料加上了一种可理解的次序。二元对立词汇间无法解决的冲突激发了神话的产生,而神话的潜在的目的正是探寻并解决结构的张力。

结构主义的主要发展是它在符号学——符号的科学——中的变形。这一发展的关键是实现了非语言系统也可以用结构主义的原则作有效的分析。批评家和哲学家罗兰·巴尔特把结构主义的原则应用于广泛不同的西方文化器物,如知名的时尚、饮食、摔跤等。巴尔特首先从研究出发孤立出一个系统,再去勾勒该系统得以运作的符号学法则,例如时尚,巴尔特认为时尚的符号学逻辑构造了一个自治的"时尚系统"。

1960年代出现的后结构主义视角,作为对结构主义文化决定论观点的部分批判,突出强调意外事件、行动和文化系统的改变。这一激烈的批判反应(reaction),首要的贡献者包括雅克·德里达(解构主义【Deconstructionism】)和米歇尔·福柯谱系学【Genealogy】,其目的在于理解语言和符号的历史建构和社会建构。安东尼·吉登斯批判地继承了后结构主义关于结构的结构化的洞见,并以此为基础,构造了他颇具影响的结构化理论。吉登斯论证了在行动赋予结构以现实性的同时,结构也限制了行动——包含在实践概念下的一个动力过程。结构主义对文本和符号的偏爱也影响了后现代主义。例如,让·鲍德里亚对于消费文化符码的探讨,就借助了结构主义研究意义的方法。

——George Ritzer
Todd Stillman
(王玥译　赵锋校)

参考文献

Barthes, R. (1967). *Elements of semiology*. London: Jonathan Cape.

Baudrillard, J. (1998). *The consumer society*. London: Sage. (Originally published in 1970)

Giddens, A. (1979). *Central problems in social theory*. Berkeley: University of California Press.

Lévi-Strauss, C. (1963). *Structural anthropology*. New York: Basic Books.

Saussure, F. de (1966). *Course in general linguistics*. New York: McGraw-Hill. (Originally published in 1915)

结构化访谈(Structured Interview)

结构化访谈是这样一种访谈,即在访谈开始之前,访谈员所要问的问题,以及多数情况下,回答者所需使用的回答选项不仅已经完全形成了固定的格式,并写在访谈表【Interview Schedule】中。结构化访谈,有时也被称作标准化访谈(standardized interviews),普遍地应用于社会调查之中,其目的在于生成关于总体的描述。

结构化访谈可以与半结构访谈【Semistructured Interview】和无结构访谈【Unstructured Interview】比照。访谈员在实施半结构访谈时,他或她有一份所要谈论的话题或问题的列表,但是问题的具体措辞则留给了访谈员。此外,访谈员还可以根据自己的意愿询问那些没有事先列在访谈提纲上的额外话题或问题。而在非结构访谈中,访谈员根据自己的意愿,在访谈过程中确定询问哪些问题,以及应当如何措辞。

如果访谈的目标在于形成统计数据,那么结构化访谈的应用就是必需的。因为除非所有的回答者回答同一个问题,否则问题的答案就不能有意义地被制成统计表。

例:您赞成或反对美国总统履行其职责的方式吗?

上面如例示的问题经常被用于形成关于总统赞成评定的统计表。人们可以计算回答了这一问题的民众样本,以及回答"赞同"的民众的百分比。赞成的百分比可以根据时间制成统计表,并通过对比总统执政的其他时间,以及其他总统的执政情况,来考察民众对总统的相对满意或不满意感。

设想一下,如果,那些民众被问的不是有关总统的,而是他们是否赞同或不赞同处理政务的方式。再设想一下,如果那些民众被问的不是他们是赞同或是不赞同,而是他们是支持还是反对总统所做的事。虽然这些措辞上的变化似乎表示民众被问了相同的问题,但是研究已经显示,在问题或回答选项的措辞上的细微变化可能会对所获得的答案产生明显的影响。简言之,如果在民众回答问题的措施上有此类变化,那么他们的答案的统计表将是不可解释的。

结构化访谈的发展

在调查研究的早期阶段,半结构访谈是规范的方法。访谈员带着一份话题列表就被派出去了,而他们可以根据自己的意愿设计他们自己的问题,引出人们的回答。基于下面三项考量导致调查方法获得了新的发展,进而使得结构化访谈普遍地被应用:

1.研究显示在问题或给予的回答选项的措施上的细微变化对人们给出的答案有明显的影响。

2.当访谈员可以根据自己的意愿确定问题的措辞时,毋庸置疑,他们会以不同的方式组织词汇,形成问题,来达成相同的目的。其结果是,调查所获得的答案常常同那个访谈员实施访谈相关。

3.当容许回答者用他们自己的话回答问题时,通常难以比较他们回答的内容。例如,若一个回答者用"不太坏"来描述她的健康状况,而另一个人用"相当好"来描述她的健康状况,那么我们能够有把握地说,第一个人的健康状况是好于、差于,还是同等于第二个人的健康状况吗?研究者发现,如果他们迫使回答者从同一组回答选项中作选择,他们就可以更加可靠地制作答案的统计表,并比较不同回答者的答案。

因此,为了确保人们给出的答案上的差异来自于他们必须作答的内容的差异,而不是基于他们被问问题的差异,或是他们用以回答问题的措辞方式的差异,现在,按照常规,研究者都会使用结构化访谈,即无论是问题的措辞,还是回答选项的措辞都有事先的设计,且对所有的回答者都一样。

用于结构化访谈的标准

就半结构和无结构的访谈而言,访谈员可以自由地根据特定的情形,以及他们对与之谈话的回答者的特征的知觉,来确定问题措辞的适宜方式。而在结构化访谈中,访谈表的设计必须形成一组"适宜"应用于所有回答者的问题和回答选项。这就意味着访谈表的设计必须予以认真细致地完成;为了有效地达成这一目标,一份结构化访谈表必须满足若干标准:

1.问题和回答选项的用词须有一致的理解,即所有回答者以同一种方式来理解。如果问题中包含有意义不明或意义模棱两可的词,就会产生质量较差的数据。这是因为,实际上,回答者回答的不是同一个问题,而是按各自对问题的理解作答。

2.应当对那些所有或几乎所有回答者都能够根据自己的知识情况予以作答的事

物提问。如果访谈员试图收集的信息是人们无法回忆的,或他们从来没有的经验,那么所得结果的质量不会很好。

3.问题的措辞需要非常确定,以致访谈员可以确切地逐词地阅读问题,而无须增加、减少或变动词语。如果访谈员为了理解问题,而不得不改变它们,那么,实际上,不同的访谈员将问出不同的问题,进而结构化访谈的主要目的也会被破坏。

4.问题的前后次序必须形成合理的方案,适宜于访谈的互动。问卷的设计必须考量问题的顺序,这样回答者才会不被重复地问同一个问题,同样,访谈者才能知道那些问题适用于或不适用于回答者。

5.当回答者从回答选项列表中选择答案作答时,回答选项必须适合问题;此即,回答选项必须给出详尽的、互斥的,包含有回答者可能想要给出的所有可能的答案的列表。

6.最后,如果一项结构化访谈准备以多种语言来实施,访谈的设计者需要思考他们所写问题和回答选项翻译成新的语言会多大程度上不同于原来的语言。

访谈步骤

结构化访谈的目的在于尽可能一致地向所有回答者询问问题。为实现这点,访谈员在实施结构化访谈时,需要遵循特定步骤。

1.他们要以问题出现在访谈表上的顺序来提问,同时,他们要确切地按照问题措辞的方式来提问。

2.当一个回答者没能完整地回答某个问题时,访谈员要询问接下来的问题(被称作非直接探询),这样就不会增加一个答案凌驾于另一个之上的可能性。

3.访谈员不加质疑地记录下回答者的答案,即如回答者所给出的那样,确切地记录。

4.访谈员需要给予最少的个人的或判断性的反馈,努力形成一种专业的关系,把提供精确的答案置于首要位置。

检验一份结构化访谈

因为一份结构化访谈必须符合如此多的标准,所以在将它应用于实际的调查之前,评估访谈表就很重要。有三种常用的评估。

1.对访谈的主题素有研究的专家可以就问题是否提的正确,以及是否涵盖了需要涵盖的材料,而提出意见。精于调查问题设计的专家常常可以就哪些问题可能产生疑惑,哪些问题难以回答,或难以用标准化的方式实施,提供有用的信息。阅读方面的专家可以评估结构化访谈表,以确认将问题放置在能使其更清楚、更简单或更易被理解的位置上。双语评估者可以确认那些可能难于译成其他语言的问题。

2.问题的认知检验在 1990 年代开始广泛应用。尽管认知检验有许多种形式,但是基本的方案是让回答者回答一些测验问题。接着,回答者被要求就他们如何理解问题,以及如何思考回答问题的方式作出解释。通过这一过程,研究者可以确认问题措辞中的模棱两可,回答者在形成问题的答案时遇到的困难,以及多大程度上答案正确地反映了回答者必须作答的内容。

3.先于一项实际的调查,对结构化访谈进行田野前测是第三种用以评估结构化访谈表的主要方式。通常,有经验的调查者会先就一批实施结构的访谈,而这批人同实际的调查参与者很相似。访谈员报告有疑问的问题是确认调查问题的疑问的一种方式。另一种方法是在回答者的容许下,录音记录下前测的访谈过程,并倾听这些录音,系统地统计那些被逐字地阅读的问题的比率,那

些回答者在回答问题时有困难的比率,以及那些回答者请求澄清的比率。

因为访谈期间,访谈员不能任意地改变问题,所以在结构化访谈表被使用之间,彻底地检验它们就尤为重要。

结论

结构化访谈的发展是调查研究和社会科学领域在20世纪中的一项重要进步。在一项有上佳实施的社会科学调查中,结构化访谈的设计和评估是最重要的要素之一。在数据收集之前,对结构化访谈中的问题,缺乏仔细认真的设计和评估可能成为调查中误差和不可信的一个重要来源。

——Floyd J. Fowler, Jr.

(赵锋译校)

参考文献

Converse, J. (1987). *Survey research in the United States*. Berkeley: University of California Press.

Fowler, F. J., & Mangione, T. W. (1990). *Standardized survey interviewing*. Newbury Park, CA: Sage.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys*. New York: Academic Press.

结构化观察(Structured Observation)

结构化观察(有时也称为系统化观察)是一种数据收集技术。它有两个确定特征:第一,它是包含范围广泛的观察技术家族的一个组成部分,其中调查员直接收集资料而不借助回答人、访谈员等中介;第二,它是结构化的或系统化的技术,即根据仔细界定的规则和事先设定的程序来收集数据。它可以在广泛的社会背景下用于行为和互动的不同方面。像任何观察技术一样,它是选择性的:研究者试图聚焦于被观察的情境中的一些与特定调查目的相关的要素。

结构化观察的直接性意味着研究者可以收集到那些调查对象难以通过其他方法——如访谈【Interview】和问卷【Questionnaire】——给予的数据。这可能是由于需要的信息的性质和细节(教师在一堂课中提问的次数,一段对话中被打断的次数)。这也可能是由于有限的语言技能(如小孩子)。直接性也与观察的方式有关,观察可以在自然的情境下发生,这种情境不是为研究目的而创造的,不会因测量而打断情境。

结构化观察的本质源于使用了观察表【Observation Schedule】,借由观察表,依据事先仔细界定好的标准记录观察,而这些标准就是变量的类别,并可用于统计分析。这些变量和类别必须界定清楚才能让不同的观察者以同样的方式使用它们,从而消除源于不同观察者的特别兴趣和认知而产生的陈述差异。从这个有限但很重要的意义上看,结构化观察可以被描述为数据收集的客观方法。就像在其他类型的研究中一样,研究焦点和工具设计都会受到研究者的兴趣和价值取向的影响。然而,数据收集过程应当独立于这些价值和兴趣,同时,变量操作化的方式还应当是开放的,且是可以审查的。当一种方法仅要求对一个基本的事实属性有低推论性的测量,且仅涉及很少的观察者的判断时,结构化观察通常是最成功的方法。高推论性测量,如那些涉及评分量表,以及联合了观察者的判断和评估的测量,不可能很容易地被操作化。

观察的时间维度

结构化观察的关键要素是数据包含了一个时间维度：观察是在真实时间里被编码的，可以根据各种时间特性来分析。调查者需要考虑四种与时间相关的信息：

1. 持续时间：一类行为发生的时间总量。

2. 频次：一类行为发生的次数。

3. 时点：在一天、一周或一个观察时间段上，一次发生所处的位置。

4. 次序：行为一个接一个的方式。

研究者对上述各类信息的需求程度决定了所用观察系统的类型。

观察者的一致性

确保所有观察者都用同样的方式使用观察表是结构观察的关键，而观察者的一致性可以通过两个观察者同时对同一事件的编码来计算。观察者一致性的最简单的量度是大小指数（S/L）。当两个观察者对同一个观察片段进行编码时，用较低的频次或持续时间除以较高的频次或持续时间。这个值可能在 1 和 0 之间。非常重要的是要注意这个指数说明了观察者对于全面的发生率的一致性，而不是他们对于个别情况的一致性；即它显示了边际的，而不是绝对的一致性。因此，它并不能确定观察表的清晰性和明确性，尽管它可以表明整个计数（但不必是共同出现的不同变量的类别）是被一致估计的。

一个绝对一致性的量度是比率一致性指数（PAI）。可以用一对观察者对个别观察的一致观察数量除以他们所做的观察总量，来得到该指数。它的可能值是在 1.0（完全一致）和 0.0（不一致）之间。PAI 是最常用的观察者一致性指数，通常建议的满意水平的值是 0.7 或 0.8。一些研究者基于如下事实校正了这个指数，即观察者可能由于机会而对某个编码取得一致，即使他们实际上是以不同的方式使用观察表。关于这些校正和观察者之间一致性的更全面的讨论可以参见克罗尔（Croll，1986）以及苏恩和阿里（Suen & Ary，1989）的论述。在长期研究的情况下，调查者可以在常规间距上重新建立观察者之间一致性的水平以确保观察者不发生偏差。

一些实践考量

结构化观察是一种消耗资源的收集数据的方法。观察的程序需要广泛的发展和前测，观察者也需要认真培训和监督。一个重要的问题是反应性【Reactivity】。大部分结构化观察的研究目的是，消除或至少是最小化观察的过程对被观察事物的任何影响。这通常包含田野工作的预备期，其时观察者已经在场，但是还没有开始收集用于分析的数据，而这就会让研究对象能够适应观察者的在场。评估反应性可能既要询问观察者，也要询问研究对象（在研究结束时），从而将研究对象对观察者的反应作为观察系统的一个类别。传统上，结构化观察已经成为笔纸（penciland-paper）技术，即观察的表单不但要现场完成，还要为后来的计算机分析（如 SPSS）作准备。近来，已经发展出使用手提计算机的编码观察，而这种观察则更易于使用专门的软件作复杂的分析（如 Noldus， Trienes， Hendriksen， Jansen， & Jansen，2000）。

——Paul Croll

（王玥译　赵锋校）

参考文献

Croll，P.（1986）. *Systematic classroom observation.* Lewes，UK：Falmer.

Noldus，L. P. J. J.，Trienes，R. J. H.，Hendriksen，

A. H. M., Jansen, H., & Jansen, R. G. (2000). The Observer Video-Pro: New software for the collection, management, and presentation of time-structured data from videotapes and digital media files. *Behavior Research Methods, Instruments & Computers*, *32*, 197-206.

Suen, H. K., & Ary, D. (1989). *Analyzing quantitative behavioural observation data*. London: Lawrence Erlbaum.

结构的、聚焦的比较(Structured, Focused Comparison)

亚历山大·乔治创造的表达"结构的、聚焦的比较"是指在定性研究和小样本研究中理论发展的一般方法(George, 1979; 也参见 George & McKeown, 1985; George & Smoke, 1974)。这个方法是"聚焦的,因为它只选择性地讨论历史个案的特定方面"(George, 1979, p. 61),它是"结构的,因为在这个历史个案中,它使用一般的问题来引导资料收集和分析"(p.62)。

结构的、聚焦的比较是为了应对乔治认为的个案研究【Case Study】的缺点而发展出来的。他认为个案研究通常缺乏清晰理论目标的引导,同时,个案研究过于强调独特性,不能为累积的知识生产提供坚实的基础。相反,结构的、聚焦的比较强调分析者围绕个案提出一系列"标准化的、一般化的问题"。关于标准化,研究者使用对照比较"从几个个案中确认获得可比较的资料"(George, 1979, p. 62)。这一方法使得研究者可以从两个或更多个案中系统化地收集可比较的变量的资料。对于一般化的问题,研究者在一个高度抽象的层面概念化变量,从而系统地测量不同个案的变量。只有当使用这些一般化的概念时,分析者才有望提出那些可以用于多个个案的问题。同时,标准化和一般化问题的使用并不意味着分析者不能问关于特定个案的具体特性的问题。

结构的、聚焦的比较更明确地被应用在国际关系领域中。乔治在他的关于国际威慑的研究中有效地使用了它,并说明国际危机中至少有三种不同特征的威慑失败模式(George & Smoke, 1974)。这些模式是通过对国际危机的实际个案的密切分析而发现的。乔治对于威慑失败的复杂原因的总结挑战了其他理论,而这些理论,包括理性威慑理论,并没有以严密的个案作为证据的基础。最近,乔治的学生和其他学者在研究中使用这些方法,包括瓦尔特(Walt, 1987)和本内特(Bennett, 1999)。

社会科学领域呼吁结构的、聚焦的比较,以及更加普遍的趋势是,脱离单一个案研究,而朝向小样本研究设计。尽管"结构化,聚焦的比较"这一表述在今天很少使用,特别是国际关系理论之外的领域,但是,专注于少量个案,为定性研究建立严格的程序的一般考量已经为学者广泛接受。当代小N方法论学者经常把他们的研究描述为,使用了"系统化的和情境化的比较",而这与乔治原来的说法是非常接近的。

——James Mahoney

(王玥译　赵锋校)

参考文献

Bennett, A. (1999). *Condemned to repetition? The rise, fall, and reprise of Soviet-Russian military interventionism, 1973-1996*. Cambridge: MIT Press.

George, A. L. (1979). Case studies and theory development: The method of structured, focused comparison. In P. G. Lauren (Ed.), *Diplomacy: New approaches in history, theory, and policy* (pp.

43-68). New York: Free Press.
George, A. L., & McKeown, T. J. (1985). Case studies and theories of organizational decision making. *Advances in Information Processing in Organizations*, *2*, 21-58.
George, A. L., & Smoke, R. (1974). *Deterrence in American foreign policy: Theory and practice*. New York: Columbia University Press.
Walt, S. M. (1987). *The origins of alliances*. Ithaca, NY: Cornell University Press.

实质显著(Substantive Significance)

实质显著指观测到的效应是否大到足够具有意义的程度。实质显著的概率之所以得到发展是因为统计显著性检验【Significance Testing】可以发现非常小的效应是显著的,即便它们小到无足轻重。实质显著(也称实际的显著)代之以关注是否一个观测到的关系【Relationship】、差值或系数大到足以具有重要性。实质显著通常靠直觉来衡量,在不同的实质领域内有不同的标准。

例如,一个思考两个百分比的差值的实质显著的方法是根据该差值可能达到的程度。使用两种不同罪犯假释程序的监狱间,在再犯罪率上的最大可能差值可能是100%。若发现再犯罪率的差值为40%,50%或60%很可能被当作实质显著的,然而,得到一个2%,3%或4%的差值很可能被当作不甚重要的,即便这一差值可以通过常规的统计显著检验。

思考实质显著的另一种方法是以均值水平为最大的可能差值作比较。例如,如果再犯罪率的均值为30%,那么最多它可以被降低30%。少于该值(0.10×0.30 = 3%)的10%的重要性可能被当作太小而无足轻重,无论它是否是统计显著的。在某些领域,人们期望研究者事先设定他们所要求的实质显著效应的量级。

统计显著【Statistical Significance】检验的发展在于检验小样本中获得的效应是否有可能是真实的。然而,当应用于大样本时,统计显著检验可以发现非常小的效应是显著的。例如,当个案【Case】的数量达到数万时,0.01%的差值(比如,一种实验条件下的12.23%对另一种条件下的12.24%)可能被发现是显著的。然而,在大多数的实质领域,这样小的差值不太可能有重要意义。因此,人们提出实质显著的概念,用以唤起人们注意这一事实,即统计显著难以满足研究要求。由此,实质显著的准则就比统计显著的概念更为紧迫,即拒绝某个可能通过了传统的显著性检验的关系的重要性。

在衡量变量间关系的重要性时,相似的逻辑同样成立。如果一个0.03的相关【Correlation】是以几千个观测值为基础的,那么它可能是统计显著的,但是在许多应用领域中,它未必大到足以被当作实质重要的。

福勒(Fowler,1985)主张防止报告一个虽然统计显著但是无足轻重的效应的防卫措施是检验一项设定有非零效应的虚无假设【Null Hypothesis】。如果虚无假设设定以实质重要为根据的最小的效应量【Effect Size】,显著性检验则可以检查观测到的效应是否显著地大于这个最小的效应量。

——Herbert F. Weisberg

(赵锋译校)

参考文献

Fowler, R. L. (1985). Testing for substantive significance in applied research by specifying non-

zero effect nullhypotheses. *Journal of Applied Psychology*, *70*, 215-218.

误差平方和(Sum of Squared Errors)

许多统计方法需要使用各种类型的平方和,其中之一是平方的误差【Error】之和,即误差平方和(SSE)。可以用于该平方和【Sum of Squares】的另一名称是残差平方和【Residual Sum of Squares】。其和常常缩写成 SSE,ESS,RSS 和 SSR,其中 SS 表示平方和,E 表示误差,R 表示残差【Residual】。

这个术语源自简单物理量的测量,比如长度或质量。当这一测量重复多次时,人们通常不能得到同先前测量相同的值。然而,人们认为被测项目具有一个固定的、真实的,但是可能未知的值,同时,一个测量值同另一个的差源于测量中的误差。

设被测项目的真值记作 μ,第 i 个测量值记作 y_i。于是,观测值同真值的差 $y_i-\mu$ 即为测量中的误差。这一差值的量可能是未知的,但是它仍旧适用于思考测量值重复变动时的误差。

这一推理逻辑也适用于回答关于某个统计模型,比如回归【Regression】或方差分析【Analysis of Variance】拟合手头数据的程度的问题。根据这样的一个模型,人们可以由自变量的观测值和被估的参数值来计算因变量的预测值。借用"误差(error)"这一术语的早期测量意义,观测值同预测值间的差值被认为度量了由模型所产生的误差。就简单回归而言,该误差项可以表示成

$$e_i = y_i - \hat{y}_i = y_i - a - bx_i$$

字母 e 表示误差,两个 y 值分别是对自变量的第 i 个观测的观测值和预测值,a 是被估截距,b 是回归线的被估斜率,而 x 值是对该个体的自变量的值。对许多统计模型都可以计算相似的误差。

接着,我们想要整体地把握这些误差的量。然而,误差的和等于零,因为在求和时,同均值的正误差和负误差会相互抵消。当然,我们可以取每个误差的绝对值,这样,所有的误差即为正并可以求和,或是使用这些绝对值的均值或中位数。数学上,绝对值不像平方那样易于操作。因此,人们选择平方每个误差,使得每项都为正数,再累加这些平方。这就是所谓的误差平方和。

误差的一个更现代的名称是残差。观测值同预测值间的差值可以被认为是由于所有其他没有包括在抽样数据中的变量的净效应,不过,这些差值并无任何"错误(wrong)"的意思。由此,代之以误差平方和,我们得到残差平方和。

——Gudmund R. Iversen

(赵锋译校)

*也可参见残差平方和【Residual Sum of Squares】。

参考文献

Cobb, G. W. (1998). *Introduction to design and analysis of experiments*. New York: Springer-Verlag.

Iversen, G. R., & Norpoth, H. (1987). *Analysis of variance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-001, 2nd ed.). Newbury Park, CA: Sage.

Lewis-Beck, M. (1980). *Applied regression: An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-022). Beverly Hills, CA: Sage.

平方和(Sum of Squares)

许多统计计算都涉及这种或那种平方和的计算。这类和主要用于两个不同的目的:度量一组数据内每个观测值间的差异程度,或是作为一个估计总体参数的工具。如其名称所意味的,平方和是一组平方的数的总和。例如,数 1,2,3 和 4 的平方和等于 $1^2+2^2+3^2+4^2=30$。

当平方和用于度量一组数据内各个观测值间的差异程度时,我们通常用每个观测值减去观测值的均值【Mean】,再取各个差值的平方,最后累加所有的平方。上面四个观测值 n 有均值 2.5,其差值分别为 $n-2.5=-1.5,-0.5,0.5$ 和 1.5。对这些差值取平方再累加,可得平方和为 5.00。因此,这就是由平方和所度量的四个观测值间差异的程度。该平方和可以除以其适当的自由度【Degrees of Freedom】,即得观测值的方差【Variance】($s^2=1.67$),以及标准差【Standard Deviation】($s=1.29$)。这两个统计量都与观测值的数量有关,而且它们可以和有不同观测值数量的其他数据组相比较。对一个因变量而言,其平方和不为零的原因是该变量受到了其他变量的影响。

平方和也用于参数【Parameter】的估计【Estimation】,以替代最大似然方法。用平方和的目的在于写出数据和未知参数的和。例如,可以认为观测值为 1,2,3 和 4 的样本的总体均值位于这些观测值中间的某一处。发现该数的一个方法是问哪一个常量值 c,可以使得平方和 $(1-c)^2+(2-c)^2+(3-c)^2+(4-c)^2$ 取得其能够达到的最小值。

当 $c=0$ 时,我们已知该平方和等于 30。但是有其他的 c 值可以给出更小的和吗?使用不同的 c 值来试验是一种可行的方法。另一种方法是使用微积分,可以发现 c 值应当等于观测值的均值,即 2.5。于是,平方和等于 5.00 且不存在其他的可以得到更小的平方和的常数值。不必奇怪,根据普通最小乘【Ordinary Least Squares】法,使用平方和可以得出样本均值即总体均值的估计值。

更为重要的是,无论方差分析还是回归分析都使用这一最小平方法来获得方差分析中效应参数和回归分析中回归系数的估计值。例如,用以计算简单回归方程的斜率和截距的公式可以通过最小化残差平方和得到:

$$\sum (y_i - a - bx_i)^2$$

这就要求根据 a 和 b 取该数据函数的导数,并设定结果表达式等于零,再求解 a 和 b。

——Gudmund R. Iversen

(赵锋译校)

参考文献

Cobb, G. W. (1998). *Introduction to design and analysis of experiments*. New York: Springer-Verlag.

Iversen, G. R., & Norpoth, H. (1987). *Analysis of variance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-001, 2nd ed.). Newbury Park, CA: Sage.

Lewis-Beck, M. (1980). *Applied regression: An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-022). Beverly Hills, CA: Sage.

累加评级量表(Summated Rating Scale)

参见里克特量表【Likert Scale】。

抑制效应(Suppression Effect)

霍斯特(Horst,1941)指出,在一个有两个预测项的情况下,一个传统的抑制效应指,通过增加一个同准则完全无关,但同其他预测项(记作 P)有关的一个预测项(记作 S),所带来的预测中某个准则(记作 C)的一个增量。预测项 S 被称为一个抑制项,因为它会抑制(或清除)预测项 P 中与准则无关的方差。具体来说,一个抑制项是"一个由于将其加入一个回归方程,而能增加另一个变量(或变量组)的预测效力的变量。仅仅对那些回归权重被增加了的变量而言,这一变量才是一个抑制项"(Conger, 1974, p. 36)。一般而言,抑制效应倾向于在预测项之间存在高度关联的时候出现。

策尔戈弗和赫尼克(Tzelgov & Henik, 1991)进一步界定了上面的定义,他们认为当另一个预测项 $P(\beta_P)$ 的标准化回归系数【Standardized Regression Coefficient】比它相应的效度【Validity】系数(r_{CP})大的时候(并假定预测项 P 的赋值方向同准则是正相关的),一个预测项才被认为是一个抑制项(S)。此即,$\beta_P>r_{CP}$,而它还可以进一步表示为 $\frac{r_{CP}-r_{CS}r_{PS}}{1-r_{PS}^2}-r_{CP}>0$ 或 $1-\frac{r_{PS}}{k}>1-r_{PS}^2$,当 $k=\frac{r_{CP}}{r_{CS}}$ 时,后一个不等式意味着抑制效应是由 r_{PS} 和两个效度系数 $\left(\frac{r_{CP}}{r_{CS}}\right)$ 的比值决定的。

文献中讨论过三种类型的抑制效应:经典抑制、负抑制和相互抑制。经典抑制的特征是抑制与准则之间不存在关系(即 $r_{CS}=0$),如表 1 显示的那样。当抑制项具有一个负的回归系数【Regression Coefficient】(即表 2 中显示的-0.07)时,负抑制即出现,尽管它同准则是正相关的(即 0.15)。相互抑制表现在,抑制项同预测项的回归系数(用它们的绝对值)高于它们各自与准则的关系(见表 3)。与前面两种抑制情况相反,相互抑制中没有清楚的特征与抑制项相关,如抑制项与准则之间的零关系,或抑制项的负回归系数。策尔戈弗和赫尼克(Tzelgov & Henik, 1991)建议,相互抑制可以更好地界定为"这些变量对于它们所抑制的变量的回归权重的效应"(p. 528)。

表 1　经典抑制的例子

	效度系数	标准化回归系数
预测项(P)	0.20	0.27
抑制项(S)	0.00	-0.13

注:$r_{PS}=0.5$,$n=400$。

表 2　负抑制的例子

	效度系数	标准化回归系数
预测项(P)	0.40	0.43
抑制项(S)	0.15	-0.07

注:$r_{PS}=0.5$,$n=400$。

表 3 相互抑制的例子

	效度系数	标准化回归系数
预测项 1	0.50	0.65
抑制项 2	0.20	-0.43

注：$r_{PS}=-0.35$，$n=400$。

所讨论的抑制效应的前提是 $\beta_P>r_{CP}$，也可以用到有三个以上变量，以及多个变量的情况，如典型相关分析【Canonical Correlation Analysis】或判别分析【Discriminant Analysis】（具体讨论参见 Tzelgov & Henik，1991）。在这些复杂的情况下，尽管可以证实有抑制效应的存在，但是研究者不能辨识哪一个预测项是抑制项，哪一个预测项是被抑制的。

与传统的智慧相反，策尔戈弗和赫尼克（Tzelgov & Henik，1991）提供了有力证据说明抑制效应频繁出现。因为预测项（或一组预测项）中与准则无关的变量在回归分析中被抑制，抑制效应增加了可预测性。如果抑制效应出现在有两个以上外生变量【Exogenous Variable】的情况下，那么这些变量的理论机制应该非常谨慎地解释。让我们使用表 2 中的双预测项的例子来说明这一点。我们选择两个预测项作例子，因为可以很容易地发现抑制。假定程序员的工作表现是准则，相应的编程技能和过去的计算机经验是预测项和抑制项。尽管过去的经验与工作表现是正相关的，负的回归权重意味着有更多计算机的经验实际上与工作表现的下降相关。由于负抑制效应的性质，很清楚过去的经验没有对工作表现产生负面影响。正是对过去经验的测量抑制了编程技能的测量中与准则无关的方差，从而人为地推高了编程技能的回归系数。

——Peter Y. Chen

Autumn D. Krauss

（王玥译　赵锋校）

参考文献

Conger, A. J. (1974). Revised definition for suppressor variables: A guide to their identification and interpretation. *Educational and Psychological Measurement*, *34*, 35-46.

Horst, P. (1941). The role of the predictor variables which are independent of the criterion. *Social Science Research Council*, *48*, 431-436.

Tzelgov, J., & Henik, A. (1991). Suppression situations in psychological research: Definitions, implications, and applications. *Psychological Bulletin*, *109*, 524-536.

抽样调查（Survey）

抽样调查是为学术、政府和商业研究收集和分析社会数据的一种广为应用的方法。

抽样调查的特征

抽样调查由两个基本元素刻画其特征（Marsh，1982）：数据形式和分析方法。

数据形式

抽样调查可以产生一组由个案分列变量（variable-by-case）的表格型数据。在表格中，通常行代表个案【Case】，列代表变量【Variable】。每个单元格含有某个个案在相关变量上的属性【Attribute】的信息。

问卷【Questionnaire】广泛地用于抽样调查，因为它们以相同的方式向每个人提问，并由此为建立一个有结构的数据组提供了一个简单而有效的途径。然而，在原则上，

数据表格中的单元格可以填入任何数字，或者任何数据收集方法所得到的信息，比如访谈、观测，以及来自记录的数据（如个人档案）。

尽管数据表格中的行通常代表人群，但是它们也可以代表不同的分析单位【Unit of Analysis】，比如年、国别或组织。无论使用的是哪一种分析单位的类型，表格中每一列（变量）上的信息反映的都是同分析单位的类型相匹配的信息（例如，特定年、国别或组织的特征）。

分析方法

抽样调查的第二项特征是分析数据的方式。调查分析的一个功能是描述一组个案的特征。一个由个案分列变量的表格，就每个个案收集了相同的信息，从而简化了统计描述的工作。不过，调查研究者通常还有意于解释【Explanation】——识别因果关系。这就要通过考察因变量【Dependent Variable】（假定的效应项）中的变差和选择一个能够解释此种变差的自变量【Independent Variable】（假定的原因项）达成目标。分析包括检验，是否自变量上的变差（如收入）可以被系统地归之于自变量上的变差（如教育水平）。尽管任何此种协变关系【Covariation Relationship】都不证明因果关系【Causality】的存在，此种协变关系却是因果关系的先决条件。

不同于实验研究，其中个案间的变差由试验者所产生，社会调查依赖于个案中的既存变差。这就意味着调查分析采取相对消极的方法来进行因果推论（Marsh，1982，p. 6）。一个调查可能提问“婚前同居增加还是减少了婚后婚姻关系的稳定性?”这就不太可能采纳实验方法，即随机地将人分配进两种实验条件（婚前同居和没有婚前同居）之一，然后再看婚后哪一种关系更有利于婚姻关系的稳定。抽样调查采取了消极的方法，先识别婚前同居和没有婚前同居的配偶，再比较他们的离异率。

抽样调查的问题是两组间婚姻离异率上的任何差值都可以归之于两组间除婚前同居以外的其他差异（如年龄、宗教背景、种族、人种、婚姻史）。调查分析试图通过统计控制或通过使用不同的多元分析技术消除其他变量的影响，来控制群组间的其他有关差异（Rosenberg，1968）。

历史

作为一种社会研究的方法，抽样调查已经从19世纪英国社会改革者的原始方法发展成一种有着广泛应用的、成熟的调查工具。早期的抽样调查由英国的社会改革者实施，而他们致力于研究受迅速工业化影响的穷人的条件。这些调查在今天可能就不会被认可为社会调查。

由于需要对1930年代大萧条时期和第二次世界大战时期的贫困，失业和消费模式有更好的认知，现代调查承受了相当大的冲击。这些巨大的社会解组导致在英国和美国实施了一系列的政府调查（如英国的战后社会调查和劳动力调查）以及主要调查研究机构的建立，如美国的国家普查局和国家健康统计中心。第二次世界大战期间，美国陆军的研究机构征召了像斯图弗尔、梅东、拉扎斯菲尔德、李克特和哥特曼作为研究者，而正是这样一些人成为了调查研究的著名人物，同时也成为了调查分析领域的伟大创新者。战后，这些人建立起以创新为主的基于大学的调查研究中心，从而发展了调查的方法，并指导它们广泛地应用于商业和政治研究。

抽样方法论的发展对调查研究的接受是关键性的。在20世纪的早期，统计学家推进了抽样理论，但是即便是出现了1936年和1948年两次总统大选预测的失败，抽

样理论对调查实践还只有很少的影响。从1950年代开始,非概率抽样【Non-Probability Sampling】的失误导致概率抽样【Probability Sampling】方法被广泛采纳。尽管结构的、面对面的访谈一直是人们偏爱的收集调查信息的方法,但是逐渐降低的回答率、上涨的成本和安全的考虑已经促进人们发展出替代的数据收集方法。这些发展包括计算机和电话的广泛普及促进了计算机辅助电话访谈(computerized assisted telephone interviews,CATI)的广泛应用、计算机辅助个人访谈(computerized assisted personal interviews,CAPI)的发展,以及最近的、基于网络的调查管理方法。

抽样调查的评估

尽管抽样调查作为一种收集有关总体描述性、事实性信息的手段已经被广泛接受,但是作为了解社会现象和获得社会过程和模式的解释的手段,还是面临着许多批评。

下面给出了其中几项较为重要的批评:

• 收集抽样调查数据的方法易于受抽样问题、有缺陷的数据收集工具和方法的影响,而产生相当大的误差。

• 由于是以离散变量的形式来收集个体的信息,抽样调查将个体个案的"完整性"割裂开来,无法把握所涉及的个体的态度和行为的意义,进而鼓励了理解社会的原子论方式——作为分离个案的集合(Blumer, 1956)。

• 社会调查依赖于社会实在的实证论【Positivism】理解,同时,因为实证论已经不再为研究者所喜爱,所以抽样调查法也连带地被谴责。

• 在横断调查中,实验控制和时间维度的缺乏意味着抽样调查研究不能建立起适当的因果链和因果机制。

尽管所有这些批评可以准确地应用于每个特定的调查,但是这些缺点是否是调查方法的固有组成仍然是值得讨论的。实际的研究已经采取了许多方法来最小化调查误差的来源(Groves, 1989)。马什(Marsh, 1982)主张对于调查的一般批评只是对拙劣调查的批评,而好的抽样调查研究能够提供对社会因果关系和意义的理解。

——David A. de Vaus

(赵锋译校)

参考文献

Blumer, H. (1956). Sociological analysis and the "variable." *American Sociological Review*, *21*, 683-690.

de Vaus, D. (Ed.). (2002). *Social surveys* (4 vols.). London: Sage.

Groves, R. M. (1989). *Survey errors and survey costs*. New York: Wiley.

Marsh, C. (1982). *The survey method: The contribution of surveys to sociological explanation*. London: Allen & Unwin.

Rosenberg, M. (1968). *The logic of survey analysis*. New York: Basic Books.

存活率分析(Survival Analysis)

在计量生物学中,存活率分析是对时间到事件(即持续期)数据的分析。

——Christopher Zorn

(赵锋译校)

* 也可参见事件史分析【Event History Analysis】。

符号互动论(Symbolic Interactionism)

符号互动论是人文研究的主要传统,强调与经验世界的亲密熟悉。借鉴了实用主义【Pragmatism】,并与后现代主义【Postmodernism】相近,符号互动论避开抽象,支持用本土的、扎根的、日常的观察来整合真理。符号互动论一词由赫伯特·布鲁默在1937年创造,并对定性研究【Qualitative Research】作出主要贡献。这一词与乔治·赫伯特·米德的自我理论以及芝加哥社会学派的研究传统(特别是布鲁默、帕克和休斯)紧密联系在一起。评论者认为在符号互动论的发展中有许多阶段,即从1890年代到1930年代典范的逐渐建立,经由其他阶段,直至最近的多样性阶段,以及同后现代质询有着强烈亲缘关系的新理论的出现。

符号互动论社会学家们之间尽管有分歧,但是他们大部分都专注于三个交织的主题,每个主题都有其方法论的意义。第一个主题认为,人类区别于其他动物的是他们高度发展的符号生产的能力,这使得他们可以产生历史、故事、文化和复杂的沟通网络。这些也是互动论学者研究的问题,因为这些意义从来不是固定和永恒不变的,而总是转变的、突生的、模糊的,所以传统的访谈或问卷研究工具可能不是最适合的研究方法。相反,符号互动论要求研究者沉入到鲜活的经验世界之中。

这就点出了第二个主题:关于改变、流动、突生和过程。生活、情境甚至社会,到处总是在演变、调整和变化。这个持续的过程让互动论者关注以下诸项研究策略:获取自我意义的策略、发展自传的策略、根据他人调整的策略、组织时间感的策略、协商秩序的策略、建构文明的策略。这是对社会世界非常积极的观点,即认为,人类不断从事于自己的事业,将许多并行的努力结合在一起,并通过这些互动来建构社会。

这就提出了第三个主题:互动。所有互动论著作的焦点,既不是个人,也不是社会本身;而是联合的行动,即通过联合的行动组织起生活,聚合成社会。它关注的是"集体行为"。它最基础的概念是自我,而自我暗含着"他者"的观念总是出现在一个生命之中:我们不可能单独有"自我"。但是它的所有核心观点和概念都强调这个社会他者,而这个社会他者总是影响个人:"个人"这个概念实际上也是通过他者构建出来的。从根本上,互动论关注的是"人们如何一起做事情"(Becker, 1986)。

与其他社会理论可以飙升到理论的天堂不同,符号互动论者停留在实地。互动论的理论可以指导研究关于社会的任何和所有的东西,尽管它发现的总是关于经验研究的问题。但是原则上,互动论可以研究和探索社会世界的任何方面。

互动论的一些分支(经常被确定为Iowa学派,以及与测量联系的例如二十陈述测验),试图让理论在操作上更严格。其他分支主张它是一种不同的科学(例如安塞尔姆·斯特劳斯的扎根理论【Grounded Theory】),寻求在保持符号互动论特色的同时,让它更加严格。还有其他分支(通常与诺曼·邓津相关的以及后现代主义)常把理论看成需要采取更加政治化和批评性的立场。

——Ken Plummer

(王玥译　赵锋校)

* 也可参见扎根理论【Grounded Theory】、人文主义【Humanism】和人文研究【Humanistic Research】、诠释的互动论【Interpretative Interactionism】、诠释主义【Interpretivism】、后现代民族志【Postmodern Ethnography】、实用主义【Pragmatism】、定性研究【Qualitative Research】、二十陈述测验【Twenty

Statements Test】。

参考文献

Becker, H. S. (1986). *Doing things together.* Evanston, IL: Northwestern University Press.

Blumer, H. (1969). *Symbolic interactionism: Perspective and method.* Englewood Cliffs, NJ: Prentice Hall.

Denzin, N. (1992). *Symbolic interactionism: The politics of interpretation* Oxford, UK: Blackwell.

Gubrium, J. F., & Holstein, J. A. (1997). *The new language of qualitative method.* New York: Oxford University Press.

对称量度(Symmetric Measures)

关联【Association】的对称量度描述了两个不区分前项(或自变量)和后项(或因变量)的变量 X 和 Y 间的关系。对称量度的例子包括皮尔逊相关系数【Pearson Correlation Coefficient】、斯皮尔曼相关系数【Spearman Correlation Coefficient】、戈德曼和克鲁斯卡尔的 γ,以及点双列相关系数,此外还包括下述五种量度。

表 1 性别(X)和成功记录(Y)的分布

例示 A

	失败(Y=0)	成功(Y=1)	总 计
女性(X=1)	20=a	5=b	25=(a+b)
男性(X=0)	10=c	15=d	25=(c+d)
总计	30=(a+c)	20=(b+d)	50=a+b+c+d

$$\phi=\frac{bc-ad}{\sqrt{(a+c)(b+d)(a+b)(c+d)}}=\frac{5\times10-20\times15}{\sqrt{30\times20\times25\times25}}=-0.41$$

例示 B

X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
1	1	1	0	1	0	1	0	1	0	0	1	0	1	0	1	0	0	0	0
1	1	1	0	1	0	1	0	1	0	0	1	0	1	0	1	0	0	0	0
1	1	1	0	1	0	1	0	1	0	0	1	0	1	0	1	0	0	0	0
1	1	1	0	1	0	1	0	1	0	0	1	0	1	0	1	0	0	0	0
1	1	1	0	1	0	1	0	1	0	0	1	0	1	0	1	0	0	0	0

$\sum XY = 5, \sum X = 25, \sum X^2 = 25, \sum Y = 20, \sum Y^2 = 20$

$$\phi = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{n}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{n}\right)}} = \frac{5 - \frac{25 \times 20}{50}}{\sqrt{25 - \frac{25^2}{50}}\sqrt{20 - \frac{20^2}{50}}} = -0.41$$

注:X=1(女性),X=0(男性);Y=1(成功),Y=0(失败)。

Phi 系数

Phi 系数(ϕ)是皮尔逊相关系数的特例。当两个变量 X 和 Y 自然地两分(比如,男/女,成功/失败,存活/死亡)时,习惯上皮尔逊相关系数就被称作 Phi 系数。因为同皮尔逊相关系数有直接的关系,所以用于皮尔逊相关的虚无假设的计算公式也可以用于 Phi 系数。

设有两个两分变量(比如,性别和成功与否),性别和成功的分布可以排列在通常称为列联表【Contingency Table】的带有四个格子的表中,如表 1 的例示 A 所示。在例示 A 中,有 20 个成功的女性,5 个失败的女性,10 个成功的男性和 15 个失败的男性。根据简化的皮尔逊相关公式,

$$\phi = \frac{bc - ad}{\sqrt{(a+c)(b+d)(a+b)(c+d)}}$$

[假定$(a+c)(b+d)(a+b)(c+d) \neq 0$],如果就变量 X 而言,女性被编码【Coded】作 1,男性编码作 0,同时,就变量 Y 而言,成功被编码作 1,失败被编码作 0,那么性别(X)和成功与否(Y)间的系数为-0.41。字母 a,b,c 和 d 代表列联表的每格中的频次。我们也可以根据传统的皮尔逊相关公式来计算 Phi 系数。如表 1 的例示 B 所示范的,性别和成功与否间的相关为-0.41。负相关说明“性别”的取值(从 0~1)越高,则“成功与否”的取值(从 1~0)越低。换言之,女性比男性更少成功。

四分相关

如果 Phi 系数的计算是基于两个人为的两分变量,而实际上它们是带双变量正态分布【Normal Distributions】的连续变量【Continuous Variable】,Phi 系数则可能低估变量间的关系。为了估计两个变量间的实际关系,而不是像人为的两分变量那样,我们就应当应用四分相关公式。例如,两个变量,睡眠(X)和学习成绩(Y)是人为地两分的,这样睡眠或者被归类为短睡眠(如果一个人睡眠时间短于 7 小时),或者被归类作长睡眠(如果一个人睡眠等于或长于 7 小时),同时,学习成绩可以归类作差(如果等级的平均分低于 2.5),或者被归类作好(如果等级的平均分等于或大于 2.5)。两个变量的频次汇总于列联表,见表 2。

表 2 睡眠(X)和学习成绩(Y)的分布

	成绩差 (Y=0)	成绩好 (Y=1)	总 计
长睡眠 (X=1)	60=a	40=b	100=(a+b)
短睡眠 (X=0)	40=c	60=d	100=(c+d)
总计	100= (a+c)	100= (b+d)	200=a+ b+c+d

Tetrachoric 相关公式可以定义为

$$r_{\text{tet}} = \frac{bc - ad}{\lambda_X \lambda_Y n^2}$$

其中,n 是总的样本【Sample】量,λ_X是标准正态分布在$\frac{a+b}{a+b+c+d}$处的纵坐标(即在变量 X 上主体取值为 1 的比例),而 λ_Y是标准正态分布在$\frac{b+d}{a+b+c+d}$处的纵坐标(即在变量 Y 上取值为 1 的比例)。根据表 2 中的数据,

四分相关可以作如下计算：

$$r_{\text{tet}} = \frac{40 \times 40 - 60 \times 60}{0.398\,942 \times 0.398\,942 \times 200^2}$$
$$= -0.31$$

同 r_{tet} 相比，基于相同数据的 ϕ 小了许多(ϕ =-0.20)。值得注意的是 r_{tet} 只是当样本量大于 400 的情况下，对某个关系的近似估值。

在下面两节中，将描述两个相关的关联量度，列联系数和 Cramér 氏 V。两个量度都不需要对变量的性质作任何假定(如正态的对偏斜的【Skewed】、连续的对离散的【Discrete】、定序的对非有序的)。尽管有这一优点，但是任何一对列联系数或 Cramér 氏 V 系数都不能相互比较，除非在两个列联表中 $\min(r,c)$ 的值相等，而 $\min(r,c)$ 指列联表中的行数(r)或列数(c)的最小数量。

列联系数

我们常常实施卡方检验【Chi-Square Test, χ^2】，以估计两个以列联表形式呈现的类别变量间是否存在着关联。例如，一位汽车经销商有意根据她的顾客数据库，检查汽车颜色的偏好(如白色、黑色、金色、蓝色)和购买者的种族(如高加索人、西班牙裔、非裔美国人)间是否存在着关联。尽管她可以应用 χ^2 检验来查看两个变量是否存在任何关联，但是 χ^2 值无法提供关于关系的量级的信息，因为不同的 χ^2 值可能表示同一关联的量级。为了解决这一限制，列联系数 C 已经有了长期的应用，即

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

其中，n 是列联表中的总的观测频次。如果仔细地考察该公式，很明显列联系数不可能趋近于单位 1，即便存在完全的关联。假定存在着完全的关联，如果列联表有相同的列数(c)和行数(r)，该系数则等于 $\sqrt{\frac{c-1}{c}}$。

Cramér 氏 V

Cramér 氏 V 被发展出来，用以解决列联系数不能取得单位 1 的限制。概念上，Cramér 氏 V 系数 $V = \sqrt{\frac{\chi^2}{n \times [\min(r,c) - 1]}}$，被视作观测的统计量同它可能的最大统计量的比值。因此，这个系数可以在 0~1 取值。属于 Cramér 氏 V 系数和列联系数的虚无假设检验可以基于独立的 χ^2 检验来实施，其自由度【Degrees of Freedom】等于 $(r-1)(c-1)$。

Kendall 氏 Tau

斯皮尔曼相关系数的一个竞争者是 Kendall 氏 tau(τ)系数，它基于个案数 n 来估计两个定序变量 X 和 Y 的关系。概念上，它估计同序对和逆序对的数量差异的比例，即

$$\tau = \frac{P - Q}{\frac{n(n-1)}{2}}$$

P 是同序对的数量，Q 是逆序对的数量。对于个案 i，有 (X_i, Y_i)；对于个案 j，有 (X_j, Y_j)；如果当 $X_i < X_j$，有 $Y_i < Y_j$，或者如果当 $X_i > X_j$，有 $Y_i > Y_j$，或者如果 $(X_i - X_j)(Y_i - Y_j) > 0$，则两个个案为同序对。相似地，如果当 $X_i > X_j$，有 $Y_i < Y_j$，或者如果当 $X_i < X_j$，有 $Y_i >$

Y_j，或者如果$(X_i-X_j)(Y_i-Y_j)<0$，(X_i, Y_i)和(X_j, Y_j)则被视作逆序对。Kendall 氏系数τ可以从-1（当$P=0$时）到1（当$Q=0$时）取值。如果变量X或变量Y上有同分秩，Kendall 氏系数τ的最大绝对值将小于1。在这种情况下，研究者可以使用 Kendall 氏τ_b系数或者 Stuart 氏τ_c系数。这两个指标的区别在于处理同分秩的方式。在陈和波波维奇（Chen & Popovich，2002）的书中有应用这三个系数的例子。

如果样本量大于10，$\tau=0$的虚无假设【Null Hypothesis】检验可以通过一个z检验来实现，即

$$z=\frac{\tau}{\frac{\sqrt{2(2n+5)}}{3\sqrt{n(n-1)}}}$$

相反，则需要执行t检验来检视τ_b和τ_c的虚无假设（Kendall & Gibbons，1990）。

——Peter Y. Chen

Autumn D. Krauss

（赵锋译校）

参考文献

Chen, P. Y., & Popovich, P. M. (2002). *Correlation: Parametric and nonparametric measures*. Thousand Oaks, CA: Sage.

Glass, G. V, & Hopkins, K. D. (1996). *Statistical methods in education and psychology* (3rd ed.). Boston: Allyn & Bacon.

Kendall, M., & Gibbons, J. D. (1990). *Rank correlation methods* (5th ed.). New York: Oxford University Press.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.

对称（Symmetry）

对称指一个分布【Distribution】的形状的一般属性，或者指一个不在自变量【Independent Variable】和因变量【Dependent Variable】间做区分的系数【Coefficient】。一个正态分布【Normal Distribution】是对称的，具有钟形形状，且在均值之上和之下有相同的数量的个案记分。一个对称的分布既不向右也不向左偏斜【Skewed】。如果一个系数在其计算过程中，不需要指明自变量和因变量，它就是对称的。皮尔逊相关系数【Pearson Correlation Coefficient】是对称的，即当估计两个变量Q和Z的相关系数时，无论Q被当成自变量，Z被当成因变量，或是相反，其结果都是同一个值。相反，Lambda（λ）系数就是非对称的【Asymmetric】，因为随着两个变量中的任意一个被当作自变量，其值都会发生变化。最后，值得指出的是，有时，对称也用以指称一个社会结构的形状，比如在“网络对称（nerwork symmetry）”的情形下。

——Michael S. Lewis-Beck

（赵锋译校）

共时的(Synchronic)

共时性描绘了在某一时点上的现象,隐含着研究者应用一个截面的、快照的或静态的视角(参见历时的【Diachronic】)。

——Tim Futing Liao

(王玥译 赵锋校)

系统误差(Systematic Error)

系统误差是对一个参数【Parameter】一贯的高估或低估。当系统误差的效应未予以说明或被忽略时,分析者可能从他们的研究中得出不正确的结论。

用以下方式,分析者可以区别系统误差和随机误差【Random Error】(Walker & Lev, 1953):假设对一个带有一个真值 μ 的特征实施了多次测量。其中一些测量会产生很大的估计值,而另一些有很小的估计值。如果测量工具是无偏的,那么正和负的误差将会相互抵消,同时,平均估计值将是正确的,且集中于 μ(图 1)。

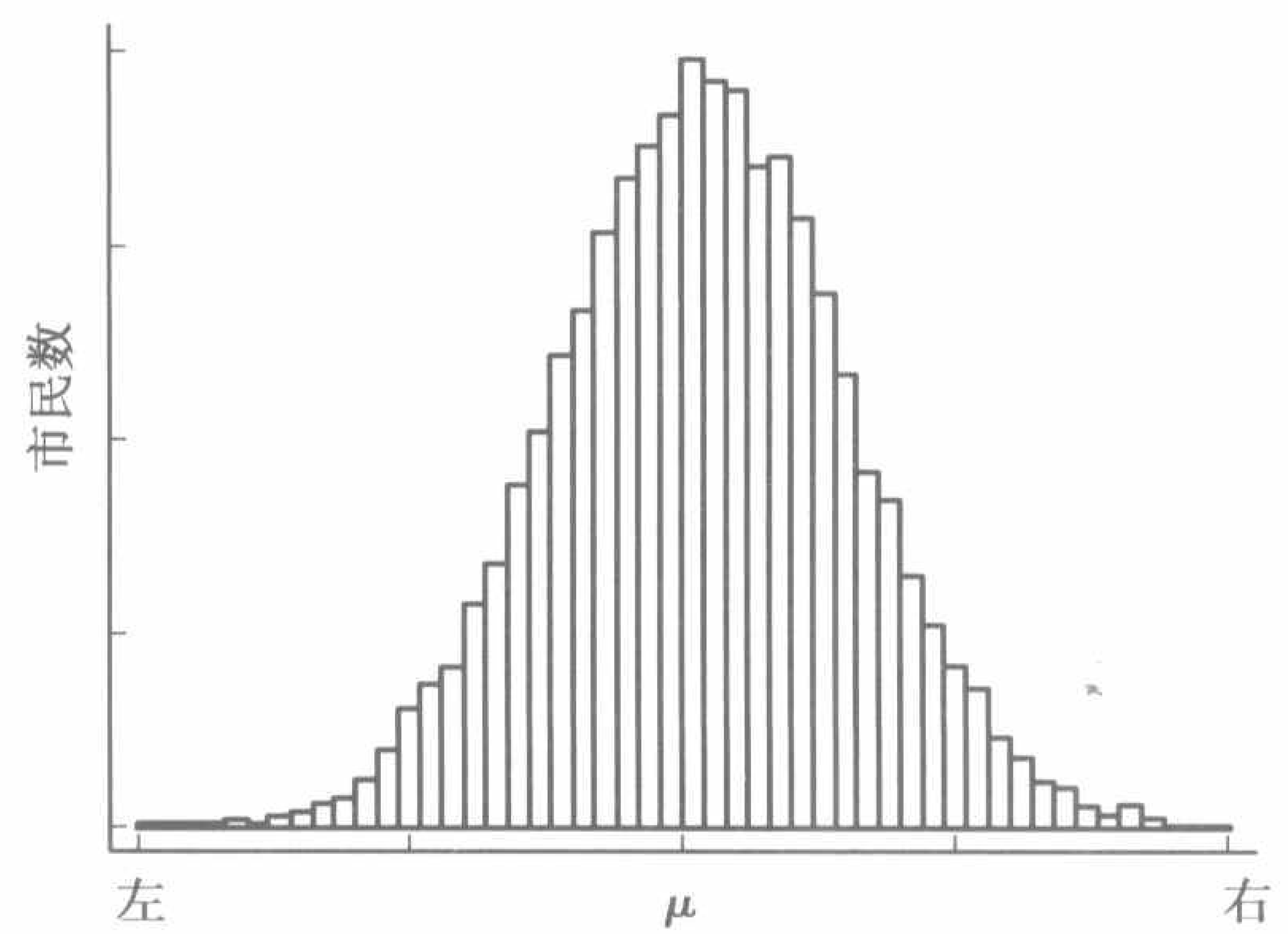


图 1 随机误差

另一方面,一个有缺陷的测量工具会给出平均而言大于(或小于)μ 的测量,从而生成系统误差(图 2)。

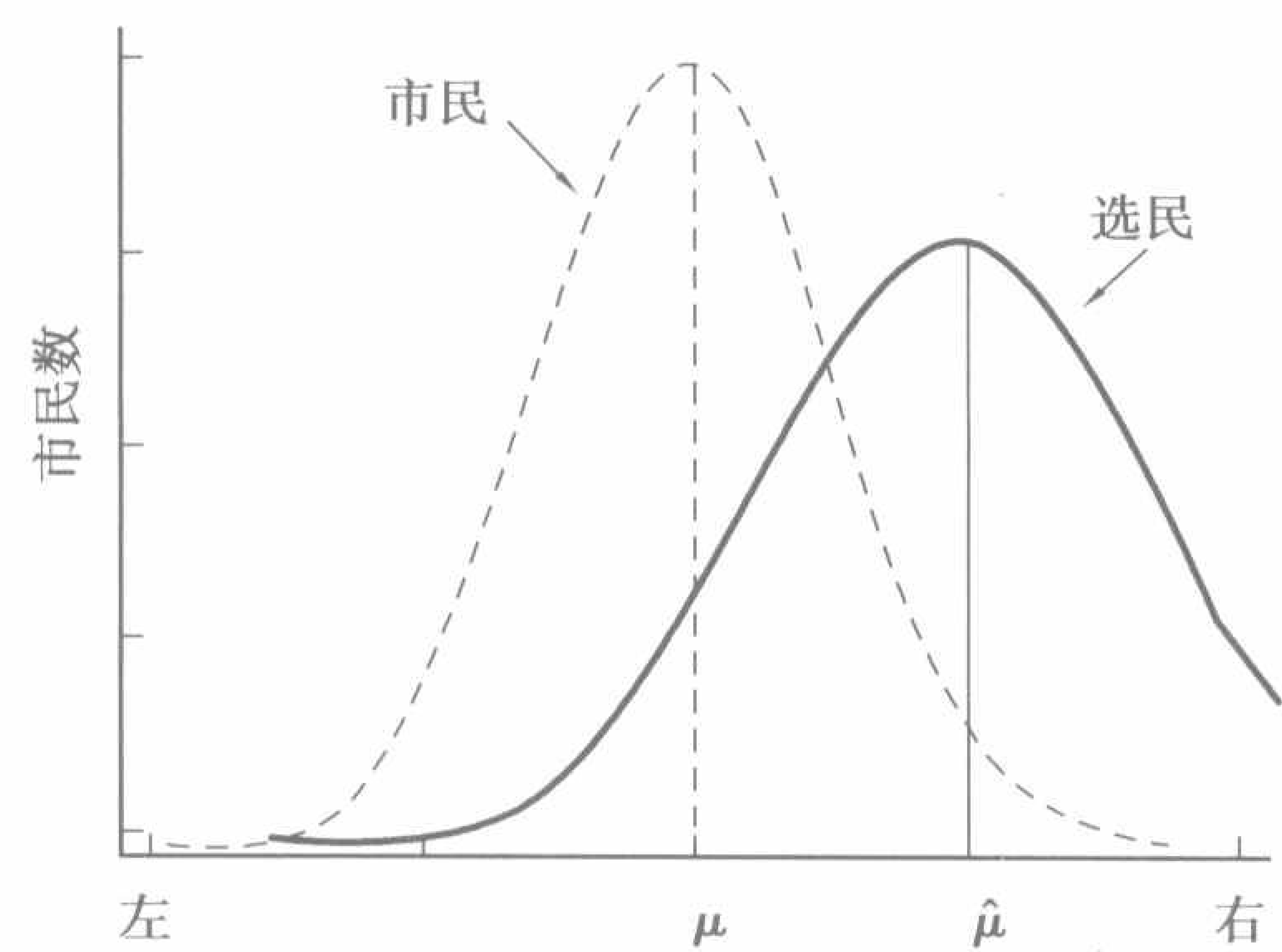


图 2 系统误差

重复测量的结果不集中于 μ,而是集中于高估了的值 $\hat{\mu}$。分析者应当注意,系统误差可能影响所有实施的测量,或部分实施的测量;也可能不同程度地影响观测值的不同子集(Nunnally & Bernstein, 1994)。在系统误差发生的每一个情形中,一个研究者的诸次测量都是有误导性的:随机误差会影响估计的精度,而系统误差使得参数估计本身是有偏的。

为了清楚地说明,考虑如下示例:在民主国家中,通过选举来测量一个总体人口的政党偏好。图 1 报告了两党制下的偏好分布:横坐标表示对特定政党的认同强度("右翼"),而纵坐标报告了特定认同水平上的市民数。μ 表示市民偏爱其他政党的比率,即"左翼"对"右翼"的比率。不过,选

举是测量一个总体人口偏好的不完美的工具。假设,偏好左翼的选民比偏好右翼的选民更少参加投票。在这种情况下,选民的分布将会类似于图 2 中的实线描绘的样子,其中均值趋向于$\hat{\mu}$(Franzese, 2002),而不像虚线的均值趋向于μ。根据这次选举,推论一个总体人口的偏好分布很可能过高地估计右翼的受欢迎程度。

这个例子显示了系统误差的一种类型,即选择性偏倚【Selection Bias】。对观测对象的选择(选民而非市民)导致了对总体人口中右翼支持率的错误计算。然而,在社会科学研究中,还会出现其他几种类型的系统误差。当使用了有缺陷的测量工具时,会产生测量误差,例如,在问卷调查研究中,回答选项的顺序的随机化是为了限制调查回答中的测量误差,因为已有研究表明,参与者的答案依赖于回答选项的呈现顺序(如 Wiseman, Moriarty, & Schafer, 1975—1976)。当两个目标变量间的关系的一个估计由于某个第三个变量(也许是未观测到的变量,或"混杂者")的排除而估计有偏时,会出现另一类系统误差,即混杂【Confounding】误差。例如,如果右翼的支持者平均地年长于那些认同于左翼的支持者,那么在年龄和政府政策支持度之间就可能显得有一个相关关系。然而,当政党认同被纳入考虑时(图 3),这就会清楚地表明,政党归属同对政府政策的态度相关联,而不是年龄。

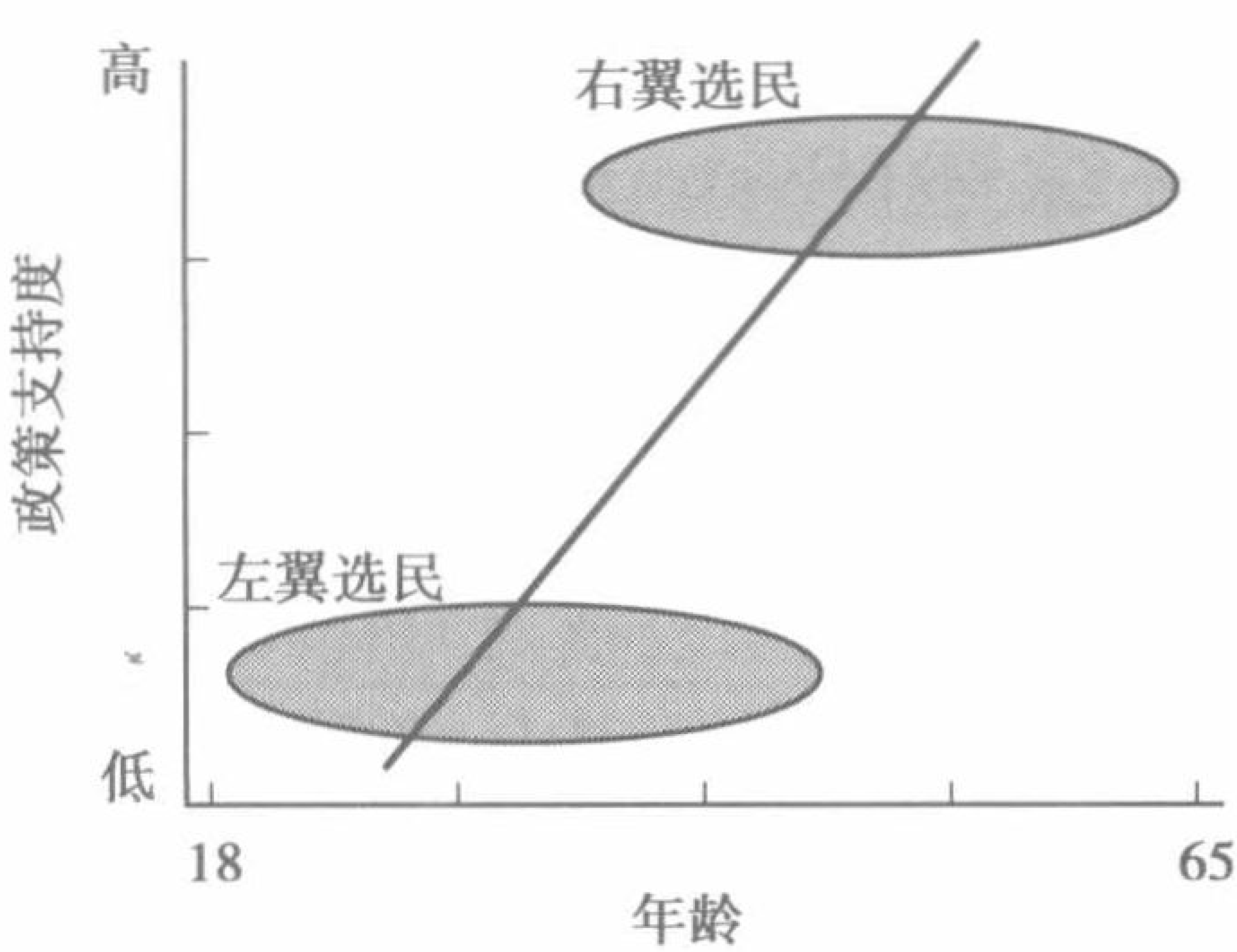


图 3 系统误差的类型:混杂误差

因此,在构建模型时,建议分析者特别要考虑不同来源的系统误差,或者在作出结论时,预先处理系统误差的影响。

——Karen J. Long

(马妍译 赵锋校)

参考文献

Franzese, R. (2002). *Macroeconomic policy of developed democracies*. Cambridge, UK: Cambridge University Press.

Nunnally, J., & Bernstein, I. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Walker, E., & Lev, J. (1953). *Statistical inference*. New York: Henry Holt.

Wiseman, F., Moriarty, M., & Schafer, M. (1975-1976). Estimating public opinion with the randomized response model. *Public Opinion Quarterly*, *39*(4), 507-513.

系统观察(Systematic Observation)

参见结构化观察【Structured Observation】。

系统综述(Systematic Review)

系统综述说明了做文献回顾的一种特别的方法。该方法要求,以明确的、可重复的和透明的过程来校核对具体问题的回答的最佳可用证据。尤其是它要求运用有力的技术查询和确认早期的研究,评估其研究质量,选取其中的研究予以评述,从研究中提取数据,以及用讲述的方式综合研究发现,同时/或者通过汇集适当的定量数据,进行元分析【Meta Analysis】来综合研究发现。系统综述的发展是为了因应传统的"叙事式"回归中出现的问题,即任意无序且受综述者个人特性影响的技术——如文献辨识的非系统方法及不拘形式的摘要——可能带来偏倚。在健康照顾研究领域,这些问题导致有效治疗手段的延迟引入以及无有效治疗手段的延续施用。系统综述逐渐从对介入手段的评估发展到对多种多样的研究问题的评估。

系统综述得益于电子数据库的发展,正是由于此发展,我们才能更广泛且更精确地搜索备选文章。重要的是,系统回顾还会寻求未发表的研究,从而减少已发表文献中存在的偏倚的风险。

系统综述批判地评估相关的前期研究的质量,有时还辅以表单和评分体系。随机化对照试验【Randomized Control Trials,RCT】常被视为用以评估介入手段的研究设计的"黄金标准",而且常常处于"证据等级"的顶端,因此,它可用以划定研究设计的等级,即根据研究设计可能产生的证据的信度。然而,由于操作的、伦理的或经济的原因,开展 RCT 并无绝对的可能性;有时,也许使用非 RCT 的证据是更适合的或者不可避免的。

一旦选好要纳入综述里的研究并提取相关的数据,就可以试着进行某些形式的综合。元分析是将来自不同研究的定量结果汇集到一起的统计方法(Egger, Davey Smith, & Altman, 2001)。有时,元分析还被当作系统综述的同义词。一般来说,元分析仅构成了系统综述的一个部分,还有许多系统综述不包括元分析。近来,人们已开始探索使用贝叶斯统计方法把元分析扩展到定性证据的可行性。更通行的做法是,以叙事的方式,而非元分析方式,综括定性证据和定量证据。

自 1980 年代系统综述推出以来,它已经成为健康科学领域中以证据为本的实践运动的核心。它是 Cochrane 协作网(一个生成健康介入的系统综述的群体的国际网络)的核心行动。最近,Campbell 协作网还将在社会的、行为的、教育的、犯罪学的介入领域开展系统综述。

好的系统综述已经产生了许多十分有益的成果。这一方法论虽然发展短暂,但未来它还会获得进一步的发展和改良。时常将某些相关数据源(如来自观察和定性研究的数据)排除于系统综述的做法已受到批评(Dixon-Woods, Fitzpatrick, & Roberts, 2001)。在近期的关于实施系统综述的指导手册中,研究者已经认识到了相关问题,同时相关的方法目前正迅速发展,以便将更多样形式的证据结合在一起。

——Mary Dixon-Woods
Alex Sutton
(赵锋译校)

参考文献

Dixon-Woods, M., Fitzpatrick, R., & Roberts, K. (2001). Including qualitative research in systematic reviews: Problems and opportunities. *Journal of Evaluation in Clinical Practice*, 7, 125-133.

Egger, M., Davey Smith, G., & Altman, D. G. (2001). *Systematic reviews in health care: Meta-analysis in context* (2nd ed.). London: BMJ Books.

NHS Centre for Reviews and Dissemination. (2001). *Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews* (Report number 4, 2nd ed.). York, UK: CRD.

系统抽样(Systematic Sampling)

系统抽样需要以固定的间距选取单位。有两种通行的应用方式:第一种方式是存在一份调查总体的单位的列表,它可以当作抽样框【Sampling Frame】来使用。随后的程序是选取列表中的每个第 A 个单位。第二种方式是不存在列表,但是可以在田野中对人流实施抽样(人流的例子包括一个博物馆的参观者,到达车站的旅客,以及跨越边界的车辆)。随后的程序是选取通过某个指定点的每个第 A 个单位。这在概念上等价于根据人群通过指定点的顺序来建立一个单位的列表,再从列表中系统地抽样【Sampling】,所以系统抽样的原则可以通过仅仅考察列表抽样来予以说明。

系统抽样的第一步是计算抽样间距,$A=\frac{N}{n}$。例如,从 $N=1\ 512$ 个单位的列表中选取一个 $n=100$ 的样本,那么 $A=15.12$。第二步是生成一个随机的起点。这步将确定列表上第一个被抽中的单位。事实上,它也将确定全部的样本,因为剩余的选项可以通过重复地累加固定间距来获得,并不再涉及任何进一步的随机机制。随机起点应当是 1 到 $1+A$ 之间的一个数字。第三步是将随机起点累加 A。因此,如果随机起点是 4.86,随后得到的数字就是:19.98,35.10,50.22,…,1 501.74。最后一步是忽略小数项——被抽中的单位是列表中的第 4 项、第 19 项、第 35 项、第 50 项……第 1 501 项。这一程序将生成确定的 n 个单位的样本。在人流抽样的情况下,N 是未知的,所以通常需要根据实际的情况来选择 A。

使用系统抽样的共同理由在于其实践上的简单性。人流抽样的情形是一个很好的例证。在许多情形下,没有其他的方法可以从一群人流中抽取一个随机样本,因为没有列表,也不可能以其他的方式造出一个列表,以便随后从中抽样。实践的考量也适用于其他情形。例如,想象一下,从放在一排文件柜的卡片文件组中抽取样本。系统抽样避免了将卡片从文件柜中移出来,再重新排序,或是给所有的卡片编号。人们所要做的只是计算所有卡片的数量。

如果被抽选的单位是以一种有意义的方式排序的,那么系统抽样具有实现隐含分层(见分层样本【Stratified Sample】)的(通常是有益的)效应。这点通常被认为是系统抽样相对于简单随机抽样【Simple Random Sampling】的一项优点。许多抽样调查在清楚的层内采用系统抽样。再次,同样的理由也适用于系统的人流抽样。如果一项抽样调查需要对一个博物馆的参观者抽样,而且,存在着一种倾向,即具有不同特征的参观者在一天的不同时段中以不同的比例出现,那么,相比当天参观者的等样本量的简单随机样本,历时一整天的系统抽样将获得更高估计精度。

——Peter Lynn

(王玥译　赵锋校)

参考文献

Levy, P. S., & Lemeshow, S. (1991). *Sampling of populations: Methods and applications*. New York: Wiley.

Scheaffer, R. L., Mendenhall, W., & Ott, L. (1990). *Elementary survey sampling*. Boston: PWS-Kent.

T

Tau(τ)

参见对称量度【Symmetric Measures】。

分类(Taxonomy)

分类是一个与生物学有特别联系的词汇,指划分类别。分类(如社会的类型、政治体系的类型)常在一些问题上产生争议,诸如划分的类别是固有的,还是由分类者制造出来的。

——Alan Bryman

(王玥译　赵锋校)

Tchebechev 不等式(Tchebechev's Inequality)

Tchebechev 有时也拼作 Chebyshev 不等式。该不等式允许在仅知一随机变量【Random Variable】的均值【Mean】和方差【Variance】的情况下,推论概率范围。这个不等式的主要作用在于以此为工具导出更丰富的理论成果(如弱的大数定律【Law of Large Number】)。

数学上,若随机变量 X,有有限均值 μ 和方差 σ^2,那么对任一 $y>0$,Tchebechev 不等式(双侧的)指出:

$$P\{|X-\mu|\geqslant y\}\leqslant\frac{\sigma^2}{y^2}$$

直觉上,Tchebechev 不等式表明,如果随机变量的方差 σ^2 较小,那么变量 X 会非常趋近均值 μ。

为说明 Tchebechev 不等式如何用以计算未知分布的随机变量的概率范围,让我们考察一个假设的例子。假如一个班级的平均分是 75,方差为 20,那么随意选择一学生,其得分低于 65 或高于 85 的概率上界为:

$$P\{|X-75|\geqslant 10\}\leqslant\frac{20}{100}=0.20$$

因为缺乏相应的分布信息以计算界限,上界必然不会特别精确。例如,假设我们已知班级分数的分布是正态分布【Normal

Distribution】,那么上述事件的精确概率近似于 0.026——同由 Tchebechev 不等式提供的 0.20 的界限有相当的差距。

定义 $y=t\sigma$,由 Tchebechev 不等式可以推出变量 X 同均值 μ 之差不大于 t 个标准误的概率:

$$P\{|X-\mu|\geqslant t\sigma\}\leqslant\frac{1}{t^2}$$

继续上面的例子,随意选择一名学生,其得分多于班级均值两个标准误的概率上界小于或等于 0.25。

对于任一 $y>0$,也可以推导出单侧的不等式:

$$P\{|X|\geqslant y\}\leqslant\frac{\sigma^2}{\sigma^2+y^2}$$

尽管 Tchebechev 不等式能用于计算未知概率分布情况下随机变量的概率界限,它最主要的作用在于作为工具推论弱的大数定律,以及用以证明均值平方的收敛隐含着概率收敛。

——Joshua D. Clinton

(赵锋译校)

参考文献

Amemiya, T. (1994). *Introduction to statistics and econometrics*. Cambridge, MA: Harvard University Press.

Rice, J. A. (1995). *Mathematical statistics and data analysis* (2nd ed.). Belmont, CA: Duxbury.

Ross, S. (1997). *A first course in probability* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

t 分布(t-Distribution)

t 分布是 t 变量的分布【Distribution】。1900 年前后,英国统计学家兼酿酒师威廉·戈塞特引入了 t 变量。他通过使用今天我们称作模拟【Simulation】的方法发现了其分布形式。他就其计算出的 750 个数据做直方图,并观察到该直方图【Histogram】不同于正态分布【Normal Distribution】。因此,他冠以新的分布 t 分布之名,并称这个变量为 t 变量。其后,罗纳德·费希尔爵士推导出此分布的数据公式。t 检验就是运用统计变量 t 而获取 p 值的统计检验。

t 变量是一变量族,变量族的任一变量有其独立的分布形态。所有 t 分布都是对称且呈钟形的,只是同正态分布相比更向两侧延展。t 变量及其分布被编号作 1,2,3,…。这数字就是变量的自由度【Degrees of Freedom】。自由度越大,分布的形态越相似于正态分布。

t 检验可用于多种不同的虚无假设。最常用的 t 检验针对单个均值【Mean】,两均值差或回归系数【Regression Coefficient】。就所有上述情形而言,检验统计量是因变量【Dependent Variable】的观测值之线性组合。

单个总体均值的 t 检验。虚无假设【Null Hypothesis】陈述 $H_0:\mu=\mu_0$。总体均值表示为 μ,假设表明总体均值等于数值 μ_0。检验统计量 t 可根据下述公式计算为:

$$t=\frac{\bar{y}-\mu_0}{\frac{s}{\sqrt{n}}}$$

此处,$\bar{y}$ 是变量的样本【Sample】均值,s 是变量的简单标准离差【Standard Deviation】,n 是样本观测值的数量。由此公

式计算所得的 t 变量值同自由度为 $n-1$ 的 t 变量分布相较。使用 t 值和自由度,可以从 t 分布表或通过统计软件查出对应的 p 值【p Value】。

当样本较小时,如少于 30 个观测值,只有变量 Y 服从正态分布,上述计算公式才有最佳结果。就大样本而言,无须此要求。

两总体均值 μ_1 和 μ_2 之差的均值检验。虚无假设陈述 $H_0:\mu_1-\mu_2=0$(或任一其他数值)。

对应 t 值具有 n_1+n_2-2 个自由度。对两均值之间的任一数值差,不仅是 0,检验都是有效的。t 统计量的公式可以在任一统计课本中找到且已被植入几乎所有的统计计算软件中。

当任一总体中变量都服从正态分布且两个分布的方差【Variance】相等时,检验效力最佳。当样本量较大,即大于 30 时,上述要求不甚重要。

回归系数 β 的 t 检验。在自变量 X 同因变量 Y 的回归研究中,我们可验证假设 $H_0:\beta=0$。对应 t 值有 $n-2$ 个自由度。当 Y 值围绕回归线呈正态分布时,检验结果最佳。

——Gudmund R. Iversen

(赵锋译校)

参考文献

Iversen, G., & Gergen, M. (1997). *Statistics: The conceptual approach*. New York: Springer-Verlag.

Moore, D. S. (1997). *Statistics: Concepts and controversies*. New York: W. H. Freeman.

Utts, J. M. (1996). *Seeing through statistics*. Belmont, CA: Duxbury.

团队研究(Team Research)

团队研究指这样一种研究,其中两人或多人充分合作,共同发表研究成果(大多数通过学术出版物),并共享发现的署名权。团队研究可能包含一个或一系列项目。它有别于单一作者的研究,也有别于有他人协助开展调查但研究的署名权归于一人的研究,因为,在团队研究中,从初始的概念化到研究发现的发表,整个研究项目的责任,在一定程度上是由合作的研究者共同承担。

团队研究可能采用多种多样的形式。最标准的形式包括两名或多名学者一起实施一个研究项目,就此研究发表一篇或多篇主题不同的论文。学者可能来自同一个机构和学术领域,或不同的机构不同的学术领域;他们可能是教师、学生或学者型专业人士;他们可能在同一地点或多个地点;他们可能有着相同的或不同的概念的和方法论的优势。

更复杂的形式包括两名或多名学者作为一个团队,在一段较长的时期中,一起作一系列的研究。即便研究系列中某些个别的研究由个别人或个别人同团队外的他人发表,他们仍可能把自己看成一个团队。

另一个更复杂的形式包括,某个实践领域的专业从业者与一名或多名学者共担责任,以研究该实践领域并在该实践领域的内部或外部发表研究发现。这种研究可能采取行动研究【Action Research】的形式,即研究同时致力于处理实践领域中特定的实践考量,同时根据问题的解决发表学术发现。

开展团队研究既有成果动机也有社会动机。成果动机包括以下信念,即多个作者的贡献(常伴随着不同类型的知识)对充分地、创造性地解决特定研究问题是必需的。社会动机包括,同朋友一起工作的需要,以

及通过把年轻学者带到研究中来帮助他们发展的需要。行动研究的动机是可以就某一实践领域的特定问题同时给予实践的和学术的贡献。

历史发展

18 世纪末期,法国最先出现学者之间的研究合作,特别是在一些硬科学和医学领域中。第一次世界大战后,这类合作大量发展。至少从 1940 年代开始,社会科学领域的团队研究,既有学术的也有行动研究的形式,已被当作一种不同的研究形式,尽管社会科学领域的团队合作要少于硬科学领域(Bayer & Smart, 1991)。

所需的特质和技能

团队合作要激发所有团队成员的动力,包括促进角色、领袖、规范、团结,以及任务和社会维度的平衡发展(Del Monte, 2000)。团队研究要想成功,需要团队成员有一些技巧和行动力。特别地,这类研究涉及相当的社交能力,尤其是处理潜在冲突和协商一致的能力。

团队成员达成一致有几个关键的方面。其中包括预期的研究成果、在个别研究或系列研究中如何划分工作和职责、要探索的研究问题、如何收集和分析资料、谁拥有收集的资料、如何解释结果的意义、资料的保密及关于在研究参与者间分享信息的决策、研究成果的发表方式、研究结果发表的署名次序。在行动研究中,需要就实现实践的和学术的目的的方式达成一致。

当有下列情况时,就以上诸方面达成一致尤其重要和困难:研究团队包括抱持有不同文化价值(例如,实践 vs.学术,着重于定性的方法 vs.定量的方法)的成员时;队员在权力、地位及关于所涉工作的类型如何做的知识上有差异时;以及/或者他们在主要目标上有差异时。而且,在整个研究过程中,可能需要在某一持续的基础上,就上述问题和获得新的一致同意保持沟通。对于那些期望不只在个别项目上实现联合研究的研究团队,这点尤为适用。

可能的好处

团队研究有几个可能的好处。一是,多种且不同立场的优势,即在其他条件相等时,一般会引出更富有创造性的研究方式。二是,相比于如果研究者自己独立做研究,他们可以处理更大的问题。三是,研究的团队方式常能取得更好的研究成果。例如,在许多领域,知名期刊中的合作署名论文的比例远高于其他期刊(Bayer & Smart, 1991)。同样,在共同署名(相对于单一署名)同论文的接受可能性及论文的引用率之间趋向一种正向关系。此外,即便团队研究可能存在困难,它也会是充满挑战并令人振奋的,至少在处理好群体协力时如此。

可能的问题

尽管有多种潜在的好处,但是在团队研究中也存在许多潜在的问题。通常,这需要解决关于署名次序或包括哪些作者的争议,以及更一般地,如何在不同成员间公平分配荣誉的争议。有时,还要考虑社会性闲散(social loafing)的问题,即研究团队的部分成员没有承担相应的职责。另外,特别是当多个学科或多种文化出现在团队中时,成员间在价值、经验、规范和不同沟通风格的差异可能引起冲突,而其中一些冲突可能无法以合作的方式来解决。如果研究团队成员之间存在权力和地位的区别,所有这些潜在问题都会恶化。如果这些问题没有得到满意的解决,团队研究的预期好处也将无法实现。

例子

实际上,两名或多名学者共同合作的团队研究的例子在所有社会科学领域无所不在。柯克曼、罗森、吉布森、泰斯卢克和麦克弗森(Kirkman, Rosen, Gibson, Tesluk, & McPherson,2003)带着其他人,展示了一个从事系列研究的团队——他们反映的是一个真正的团队。巴图内克和路易斯(Bartunek & Louis,1996)给出了许多在教育、社区环境及工作组织的内部成员与外部的研究者合作研究的例子。此类合作可能含有行动研究的成分,但这并不是必有的。

——Jean M. Bartunek

(王玥译　赵锋校)

参考文献

Bartunek, J. M., & Louis, M. R. (1996). *Insider-outsider team research*. Thousand Oaks, CA: Sage.

Bayer, A. E., & Smart, J. C. (1991). Career publication patterns and collaborative "styles" in American academic science. *Journal of Higher Education*, *62*, 613-636.

Del Monte, K. (2000). Partners in inquiry: Ethical challenges in team research. *International Social Science Review*, *75*, 3-14.

Kirkman, B. L., Rosen, B., Gibson, C. B., Tesluk, P. E., & McPherson, S. O. (2003). Five challenges to virtual team success: Lessons from Sabre, Inc. *Academy of Management Executives*, *16*(3), 67-79.

电话调查(Telephone Survey)

通过电话实施的抽样调查在美国和其他国家仍旧是最常用的询问一组标准化问题的技术,其目的在于就出于各种目的几乎任何主题,从无论大还是小的总体【Populations】中收集信息。学术组织、政府机构和商业组织常规地出于市场营销、政策评估、媒体研究、公共舆论、价值评判、人口背景、影响评估以及许多其他目的,实施电话调查。电话调查已经取代面对面的入户调查成为最常用的抽样调查技术,因为它具有改善了的概率抽样技术;强化了人口覆盖率,实践上可以覆盖每个有电话的家庭;经过测试和检验的问题措辞和问卷建构技术;改善了数据收集过程的质量控制;以及电话技术本身的进步。这些改善,配合以调查研究的消费者对于即刻回馈调查结果的需求,电话研究的较低的成本,以及面对面调查中逐渐降低的回答率,使得通过电话实施的抽样调查成为有显著优势的调查技术。

优点

几项技术有利于电话调查的流行。第一,电话的应用受到某些社会规范的强化,即它要求一个人在电话铃响3~4声内作出应答,同时,打电话的人(也就是,访谈员)也是结束谈话的人。第二,因为有更低的田野和人员成本,所以电话调查的成本低于面对面实施的调查,同邮寄调查的成本大致相当。第三,完成一项电话调查,将调查结果给予研究者或发起者所用的时间相对于其他类型的调查而言相当地少。事实上,"即时民意测验"——经常为政客所使用,其目的在于寻求对某个政策提案的公众反应——可以在24小时内给出第一轮访谈的结果。第四,因为电话访谈员比起其他调查中的访谈员更少能够了解回答者的身份信息,通常只有一个电话号码,所以电话调查加强了回答者身份及其应答内容的保密

【Confidentiality】。第五,访谈员的特征对于应答的效应要低于面对面访谈员的,因为在电话调查中,只有声音的联系,访谈员和回答者的人种和年龄一般是不可识别的。因此,由于这些物理上的特点,电话调查的回答者较少受社会期许的应答模式的影响。第六,电话调查,特别是随着实施计算机辅助电话访谈(computer-assisted telephone interviewing,CATI)的电话调查中心的设立,为整个数据收集过程的质量控制提供了机会,因为督导员可以给予访谈员直接的反馈。此外,这些设备几乎总是有能力通过一个卫星计算机终端监控访谈。第七,覆盖率不再成为问题,因为实际上100%的家庭拥有电话机,同时,随机数拨号【Random-digit Dialing,RDD】程序使得接通每个电话号码成为可能,而无须依赖一份列出地址或姓名的抽样框。针对家庭内抽样的现有技术,如最后一个/第一个生日法或家庭成员枚举法,都可以持续地实现回答者的概率选取。第八,电话调查可能就几乎任何给定的话题,敏感的或不敏感的,询问问题。因为电话已经成为一种标准的沟通手段,而且多数国家都可以具有“访谈社会”的特征,即访谈,通常由电话进行的访谈,是普遍的调查机制,所以回答者惯于向他人,即便是他们不认识的人,提供信息。

缺点

尽管有许多因素促进了电话调查的使用,但是仍然存在着不利的因素。第一,比起将某人拒之于门外,挂掉一通电话更为容易。因此,更容易的拒访使得电话调查的回答率略低于面对面调查中的回答率。改变拒访是可能的,但是这既耗时,又昂贵,而且只能产生15%~20%的成功率。第二,一次电话访谈的时间受限于20~25分钟,而面对面访谈通常可以进行更长时间,因此也能问得更深入、更细致。尽管在电话调查中有可能询问开放式问题【Open-Ended Question】,但是重新给应答编码,以及探询更深入的内容会更加困难。因此,电话调查的问卷倾向于使用结构化的封闭式问题【Closed-Ended Questions】,因为与开放式的、无结构的题项相比,这类问题看起来更易于通过电话作答。第三,因为访谈员和回答者没有身体上的显现,因此电话访谈员不能观察到被访者的周围环境,或是获得非口头的线索,而只能依赖声音的特性。电话访谈员也不能像面对面的访谈员那样有可能使用视觉辅助工具。第四,电话技术的发展曾经确实使得实施调查变得更加容易,但是最近的创新则将这项工作变得更加困难。几乎所有的家庭都拥有应答机,或来电等待,或来电显示,或是用调制解调器接入家庭线,而这其中的任何一项都会成为达至回答者的障碍。预想中的回答者几乎从不会根据电话答录机上的留言,回电给访谈员,或者是中断一通电话交谈,接听来自调查机构的来电。在这些条件下,重复地去电会增加总的调查成本。第五,通过电话的合法调查受到公众对调查一般的负面评价的削弱,还受到令人厌烦的电话市场营销或通话后的上门推销(利用调查想法获得“入户”的机会,以便贩卖产品)的削弱。这也被称作“诱售(sugging)”(以市场研究作幌子的销售)。以合法的民意测验作幌子的强制投票或政治宣传,以及“诱募”,即以调查作幌子的基金筹募,也都助长了此种氛围,鼓励了对任何访问(包含最合法的调查)的习惯性拒绝。

抽样

RDD抽样技术的发展使得将电话接入某个地区或社区的每一个拥有电话的家庭成为可能,无论他们的电话号码是否被列在

地区的电话号码簿上。因为,在某些社区,最多可能有40%的家庭电话号码或是因为家庭的选择,或是因为号码还没有进入到印制的号码簿中,而没有被列在表中,从而必须找到能克服未列表问题的方法。RDD依赖交换机使用的号码清单。因此,可能知道哪台交换机(如735,542)在某地区处于使用状态,同时也能知道每台交换机在用的四位号码的范围。每台交换机仅可能有10 000个号码。使用一个随机号码程序,就可以生成和呼叫适当的号码集。但是,这种方法的一个问题是,它没有能力摒除商业号码或移动电话的号码,从而使得接通一户合格的家庭更加困难。同时,许多家庭现在拥有不止一部电话的事实也使得这些回答者的选中概率大于仅有一部电话的家庭的选中概率。最后,尽管交换机在一个城市或地区有地理上的分布,它们却不能配置到更小的住宅项目或村社。在更小地理区域内找出符合调查资格的家庭使得抽样更加困难和昂贵,即便是能够从商业抽样卖家那里买到为顾客定制的样本。

结论

问题的顺序、问题的措辞、应答的模式、访谈关系的确立、问题的复杂程度,以及编码的问题,对所有类型的抽样调查都是类似的。对上述这些,以及其他类似问题已有相当数量的研究,而这些研究指出经过检验的关于问卷构造和问题写作的技术也适用于电话调查,而且还有利于减少问题回答过程中可归之于这些因素的变差。就电话调查而言,数据质量与其他类型的调查相比,处于同等水平上,甚至由于减少了可归于访谈员的身体在场的“需求”因素而有所提高。如前述所指出的,访谈过程的质量控制,和最终的数据质量,会因为使用CATI系统,而提高,因为系统会将问卷以一种自动的、受控的提问顺序呈现给访谈员,从而减少访谈员发生错误的可能性。由于CATI系统受到青睐,笔加纸(pencil-and-paper)式的电话访谈几乎已经消失。尽管CATI系统很昂贵,但是它们几乎可以给出即时结果,此外还强化了根据一份随机时间表安排再呼叫的能力,以及控制样本池的分布能力。目前,电话调查正被用作“混合模式”研究的组成部分——其中几种数据收集技术被用于同一个研究项目。这就有可能增加回答率和整个样本的覆盖率。现在,在世界各国中,实施调查都是大生意,而通过电话的调查则是该项事业中显著的组成。

——James H. Frey

(王玥译　赵锋校)

*也可参见计算机辅助个人访谈【Computer-Assisted Personal Interviewing】。

参考文献

Bickman, L., & Rog, D. J. (Eds.). (1998). *Handbook of applied social research methods.* Thousand Oaks, CA: Sage.

Frey, J. H. (1989). *Survey research by telephone* (2nd ed.). Newbury Park, CA: Sage.

Frey, J. H., & Oishi, S. M. (1995). *How to conduct interviews by telephone and in person.* Thousand Oaks, CA: Sage.

Lavrakas, P. J. (1993). *Telephone survey methods* (2nd ed.). Newbury Park, CA: Sage.

Lavrakas, P. J. (1998). Methods for sampling and interviewing in telephone surveys. In L. Bickman & D. J. Rog (Eds.), *Handbook of applied social research methods* (pp. 429-472). Thousand Oaks, CA: Sage.

证词（Testimonio）

证词一般被界定为对于有社会意义的经验的第一人称的叙述，其中叙述的声音来自一位典型的或特别的目击者或主角，这些人把他们自己当作了他人的代表，而这些被代表者虽然曾经历过相似的生活处境，但是很少与其书面表达。因此，证词是"非文艺的文学（literature of the nonliterary）"，既包括电子复制品——通常辅之以一位访谈者/编辑的声音，也包括创造性地重新安排历史事件，使其有典型性和"真实"性的印象，也使其常常映射出社会转型。在1970年代和1980年代，出现许多证词，特别是在整个拉丁美洲，包括了本土代表、劳动者和游击队领袖的叙事。在1990年代，证词转向非政府组织和各类社会运动的代表，他们既支持人权、女性权和同性恋权，也支持生态运动和反全球化运动。《我，里戈韦塔门楚》（*I, Rigoberta Menchú*）成为许多证词中最著名的一个，而最近对这一文本的抨击已使得作为一种可行的表达形式的证词成为有问题的概念。

证词暗指有证明作用或支持作用的宗教或法律的证言，同时，整个叙事单元通常就是一种生活或一段有意义的生活经历。许多评论已经将其视为那些被资本主义现代化所边缘化的人一种表达形式。证词也被看成是有教养的知识分子的介入或调停。埃尔兹别塔·斯克洛多夫斯卡（Sklodowska，1992）以一种文学的方式给予证词最细致的分析，约翰·贝弗利（Beverley，1989；Beverley & Zimmerman，1990）则呈现了对证词的社会维度最有影响的阐释。对贝弗利和齐默尔曼（Beverley & Zimmerman，1990，p.173）而言，"证词通常是由主角或目击者对事件的详述而形成的中篇小说长度的第一人称的叙事"。此外，证词既能映射出社会和文学形态的整体转型，也能映射出跨国生产模式的整体转型。

请注意，证词类的文本长期处于文学的边缘地位，贝弗利（Beverley，1989）论证说，证词的突出特征在于它同已经建立起的文学的、美学的标准，以及文学自身的规范体系的冲突关系。证词于是可以被定义为"非小说的、大众—民主式的传奇叙事，因为叙事有一种转喻的功能——意味着任何被叙述的生活可以有一种代表"，能够唤起一个含蓄的"关于其他可能的声音、生活和经历的多声部合唱"（Beverley & Zimmerman，1990，p. 175）。

证词中呈现了"真正的（real）"大众声音，在不同程度上，是一种虚幻的或不确定的"现实效应（reality effect）"——很容易就能被推翻（参考大卫·斯托尔对里戈韦塔·图姆其人和其书的质疑，Stoll 1999），但是效应的持存却有助于解释证词的特别力量。一份证词文本同时代表了某些被假定为真的视角和问题，但它从不是"真相（truth）"。

近年来，伴随着产生关键事例的社会动荡，证词已经不流行了。然而，作为性别和其他身份—外形差别的一种表达方式，证词，无论其自身还是它在小说、诗歌和戏剧中的使用，仍将作为一种可以为那些同底层社会和社会运动相联系的个体所用的、漫无边际的模式，长期地保留在拉丁美洲和全球文学中。

——Marc Zimmerman

（王玥译　赵锋校）

参考文献

Beverley, J. (1989). The margin at the center: On testimonio. *Modern Fiction Studies*, *35*(1), 11-28.

Beverley, J., & Zimmerman, M. (1990). *Literature and politics in the Central American revolutions.* Austin: University of Texas Press.

Menchú, R. (1984). *I, Rigoberta Menchú: An Indian woman in Guatemala.* Ed. and intro. Elisabeth Burgos-Debray. Trans. Ann Wright. London: Verso.

Sklodowska, E. (1992). *Testimonio hispanoamericano: Historia, teoría, poética.* New York: Peter Lang.

Stoll, D. (1999). *Rigoberta Menchú and the story of all poor Guatemalans.* Boulder, CO: Westview.

Zimmerman, M. (2001-2002). Rigoberta Menchú, David Stoll, subaltern narrative and testimonial truth: A personal testimony. *Antípodas: Journal of Hispanic and Galician Studies*, *13/14*, 103-124.

再测信度(Test-Retest Reliability)

在一个测量上,由于回答人状态的暂时变化,个体回答或观测得分趋向于波动。再测信度给出了对同一量具历时响应的稳定性(或不稳定性)的一个估计。换言之,它估计了在整个逝去的时间周期内,由于某些因素(如成熟、记忆和紧张事件)的出现而导致的测量误差【Error】的量。估计【Estimation】再测信度的步骤如下。在两个间隔一定时间长度的不同时点上,对同一组参与者实施同样的测量。再用以下的公式,计算量得的两组得分的相关系数:

$$r_{XX'} = \frac{\sum (X_i - \bar{X})(X'_i - \bar{X}')}{ns_X s_{X'}}$$

其中,X 和 X' 分别代表同一测量的两次执行,s_X 和 $s_{X'}$ 是 X 和 X' 的标准差【Standard Deviation】,n 是样本中回答人的数量。这个相关【Correlation】系数代表了对再测信度的估计,也被称作稳定性系数。

理论上,再测信度估计值的范围是 0~1,尽管由于测量受波动的反应影响,也可能观测到负值。一般而言,并不存在能够表明一个再测信度具有可接受水平的公认的最小值。此外,再测信度的期望大小可能依赖于被评估概念的性质。例如,就一个用以估计短暂特征(如心情)的测量而言,期望有很高的再测信度是不现实的。因此,相比于态度测量或民意测验,对那些被认为用以估计更稳定构念的能力测验或人格测量而言,再测信度倾向于有更高的信度。

测验与再测方法是一个恰当的信度估计技术,当:(a)在两次实施的间隔中,被估计的特征没有变化或者变化很小时;(b)测验内容不会影响对随后呈现的刺激的响应。关于第一种情形,两次实施间逝去时间的量是一个重要但常被忽视的问题。时间间隔的量应当足够长,以避免任何来自于记忆或实践的效应,但是,不应当长得以至于参与者身上令人感兴趣的特征的水平实际已经变化了。提示读者,两次测量实施的时间间隔的确定应当以相应构念的理论为指导。根据智力发展理论,在 20~25 岁,成人的智力得分不会出现大幅变化。以 5 年为时间间隔的高再测信度(如 0.80)相较于以 2 周为时间间隔的同样再测信度,前者给出的证据更有效【Validity】。另一方面,就婴儿而言,如果测量时间间隔期长于一年,一项可能可信的智力测量就可能有很小的再测信度。

关于第二种情形,相关研究已经表明,一项再测信度估计可能受到心理因素的影响,如认知测试的练习效应或态度测量的延续效应。例如,参加同样认知能力测验的人可能由于第二次阅读而更好地理解了问题,或者,人们仅仅因为记住了

他们对先前题目的回答，而可能对调查问题有相当一致的反应。这些偏倚【Biase】都可能膨胀信度估计值，比如当参与者以一种虚假一致的方式做答时；但是，它们也可能导致一个更低的再测信度估计值，即如当参与者在一项认知能力测验的第二次测验中得分更高，从而导致两次测验得分间缺少一致性时。

——Autumn D. Krauss
Peter Y. Chen
（马妍译　赵锋校）

参考文献

Crocker, L., & Algina, J. (1986). *Introduction to classical & modern test theory*. New York: Harcourt Brace Jovanovich.

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 105-146). New York: Macmillan.

Pedhazer, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Lawrence Erlbaum.

四分相关(Tetrachoric Correlation)

参见对称量度【Symmetric Measures】。

文本(Text)

术语文本是一个魅惑的词，其意义已经极度宽泛。历史上，最初它同书面的沟通形式联系在一起，特别是那些来自圣经和法律的原始材料，伴有一种权威性和权威意义的意思。这一意思转化成一个作者的观念，即作者被认为是某一文本之意义的唯一根据。因此，以文本为本的批评将其基本理论根据立基于确定和注解作者意图的意义。此种文本概念已经在一些主要方面被彻底革新了。当下的问题不仅是作者意图的意义是否可知，而且还在于我们是否能够谈论某个文本的唯一的、最终的或“真正的”意义。读者要从文本中解读出他们自己的意义，而这些意义总是背景—依赖的。在一个极端上，这导向了历史主义和文化相对主义【Relativism】的还原，但是拒绝狭隘地专注于对著作权的研究，以及拒绝文本的单一意义的观念则是有充分理由的。文本，无论是在一段时期内，还是超出了一段时期之外，其意义都是多重的和可变的。正如一个文本是由特定的符码和惯例所构建，它也由此而得到阐释和理解；这就是它的内在的语义不稳定性(semantic instability)。一个文本当然不可能仅仅意味我们给予其的意义，尽管如此，我们总是必须从呈现在页面上的内容，获得我们的阐释线索。如今，在这个框架下，它不仅是书面的或由词语进行的沟通，文本这一术语已然拓展到如此多样的文化形式中，诸如电视、电影、摄影、流行歌曲和音乐、体育和时尚、广告、家用电器和建筑。完全地关注于上述文化形式的文本性会导致形式主义的和反实在主义的立场，如某些后结构主义【Poststructuralism】的分支，即认为文本处在一个无尽的意义生成和互文性的游戏之中，除了其自身和其他文本，无现实的参照。更为普遍接受的观点是，文本同它们的社会、文化和历史的背景之间的关系，由叙事、类型(genres)及言语所中介，

即它们不仅运作于这些中介之中,而且还通过这些中介获得它们的广泛意涵。虽然所有主要的人文和社会科学学科当下都不得不同上述观点的多种含义相妥协,但是,仅仅或主要地基于文本分析(textual analysis)的阐释和理解的适当性已经受到了普遍的质疑。当我们参加到不同形式的沟通之中时——它们可能通过页面、屏幕或扬声器传达给我们,为了获得对给予我们的意义的广泛理解,分析文本不可或缺,但是,这样做并不能告诉我们包含在文本的生产和消费中的所有内容。除了文本分析,广泛的理解还需要其他的研究方法。自我限定式文本分析的危险在于假定,它能够涵盖有益于一个文本的建构和阐释的所有事项,似乎所有的东西真的可以仅从文本中"读出来"。这引起了进一步的危险。声称一个文本,就广义而言,是结构的和聚合的,中介了它所说的世界,这是一回事;但是无论何时何地,断言这个世界仅仅建立在文本和文本的相似品之中,就是另一回事了。于是,危险在于把世界文本化。这意味着,仅仅依据文本来看世界,以为文本之外无物存在。相反——虽然我们仅仅在文化文本中将这个世界变得有意义,但是,它们的意义在于,它们为这个世界建立了参照,而文本之外,世界存在。

——Michael J. Pickering

(王玥译　赵锋校)

参考文献

Hoey, M. (2001). *Textual interaction*. London: Routledge.

Lehtonen, M. (2000). *The cultural analysis of texts*. London: Sage.

Titscher, S., Meyer, M., Wodak, R., & Vetter, R. (2000). *Methods of text and discourse analysis*. London: Sage.

理论抽样(Theoretical Sampling)

资料和理论之间的关系位于经验社会科学的中心,抽样【Sampling】则有助于界定这种关系。随机抽样【Random Sampling】生成的数据【Data】能够验证既有理论【Theory】的假设。相反,包含在扎根理论【Grounded Theory】方法(Glaser & Strauss, 1967)中的理论抽样是一种分析过程,即决定下一步收集何种资料,以及应当在何处找到它。它是分析式归纳【Analytic Induction】的一种变体,其重点在于,为了发展理论,使理论系统地从资料中浮现出来,分析者将搜集、编码【Coding】和分析资料的工作结合在一起。

此过程的中心问题集中于搜集新资料的理论目的,以及决定为此目的需要何种资料。这个过程同持续比较法【Constant Comparison】方法一同进行,而且这一过程要求研究者在整个研究期间持续地追索类别、情境和分析的维度。通过结合理论抽样和持续比较,可以更有效地发展出相当完整且受支持资料约束的理论。

一个正在发展的理论的综合可以由几种方式获得,但是重点在于两个主要的机制。第一个机制是选择搜集新资料的地点,而这些地点具体地表达了基于数据所做的暂时假设。马伊内斯(Maines,1992)在纽约地铁乘客的研究中展示了这一过程。为了解释乘客的分布,他先从地铁车厢中的乘客搜集观察资料,进而引出的假设建议应该从在地铁月台上等地铁的乘客中搜集观察资料,又从此引出新的假设,建议从在地铁车厢内移动的乘客中搜集观察数据。最后

引出的假设要求比较两类乘客的访谈资料：一类乘客进车厢后就待在车厢中；另一类是走到别的车厢找座位的乘客。在研究过程的每一个阶段，研究者构建其暂时的理论，以解释现存的资料；同时，那些暂时理论的性质提示出，为改善正在发展的理论的解释力所需新资料的种类和来源。

第二个机制是编码【Coding】。斯特劳斯(Strauss, 1987, Chap. 3)确认了三种与理论发展相关的编码程序。第一个是开放编码(open coding)，即为了发现类别和中心主题而去探究资料的过程。这一过程通常出现在研究过程的早期。第二个是主轴编码(axial coding)，着重于核心类别及其属性的辨识，以及提出相关的理论，说明属性(或维度)出现的条件，以及它们与其他类别和属性的关系。第三个是选择编码(selective coding)，即就类别和子类别间的关系，做出更具理论目的的专门说明。

苏利耶、布里特和马伊内斯(Soulliere, Britt, & Maines, 2001)示范了，怎样结合起理论抽样和对核心类别的编码使其与概念建模相兼容。通过专注于"专业化(professionalism)"这一类别，他们展示了如何在新资料的基础上，重新设立一个概念—指标项模型，以及如何将其作为一个概念—概念模型呈现出来。这一建模过程是对于扎根理论宗旨的一个清楚的应用，而扎根理论的宗旨是将理论从具体推向更抽象和更一般的系统表述。

——David R. Maines

(王玥译　赵锋校)

参考文献

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory*. Chicago: Aldine.

Maines, D. R. (1992). Theorizing movement in an urban transportation system by use of the constant comparative method in field research. *Social Science Journal*, *29*, 283-292.

Soulliere, D., Britt, D. W., & Maines, D. R. (2001). Conceptual modeling as a toolbox for grounded theorists. *Sociological Quarterly*, *42*, 233-251.

Strauss, A. (1987). *Qualitative analysis for social scientists*. New York: Cambridge University Press.

理论饱和(Theoretical Saturation)

理论饱和是定性资料分析的一个阶段，即在这个阶段上，研究者持续地抽样和分析资料，直到没有新的资料出现，而且理论中所有的概念都很饱满。形成理论的概念及概念间的联系已得到验证，且不再需要其他资料。理论的所有方面都不再是假设性的。所有概念的界限都是明确的，联合的概念都已被识别和勾勒出来。反面案例【Negative Cases】必定已被辨识、被验证、被充实并被纳入理论框架中。

理论饱和的基础是理论敏感性(theoretical sensitivity)概念。理论敏感性的主要假设是，资料分析是资料驱动的(data driven)。此即，除非类别"为自己争取到了"进入理论框架的价值，它们将不会出现(Glaser, 2002)。这意味着，某些总是存在于研究中的类别，如性别，将不会在研究中取得一定的位置，直到资料分析显示这一概念需要被纳入理论框架。一旦这些概念出现，抽样需要持续进行，直至这些概念实现饱和。

理论饱和这个术语主要使用于同扎根

理论【Grounded Theory】相连的研究，但也可以应用到任何以发展一个定性驱动的理论（qualitatively derived theory）为最终目标的定性研究中。定性驱动的理论有别于用于定性研究的理论——如果说定量研究中的理论基于并得自先前的质询（而目前的项目是此前工作计划的一个延伸），那么定性质询中的理论则是从资料中系统地发展来的。

尽管结果仍然是一个理论（在其中它是对资料的概念抽象），定性驱动的理论与定量调查中的理论区别在于，前者的理论更贴近现实。定性驱动的理论是质询的产品或结果，而不是待检验的一个概念框架，如其在定量质询中一样。在定性驱动的理论中，概念必须得到充分的发展，而且它们与其他概念的连接必须被清楚地描述（Morse, 1997）。用以评估突显的理论的准则是，它必须是可理解的、俭省的及完整的；它必须有把手；它必须是有用的（Glaser, 1978）。

应用理论饱和发展一个理论的一个例子是，斯特劳斯和他的同事的经典专著。在这个扎根理论的研究中，他们描述了患慢性病的个体经验的诸多方面，包括生病和就医的轨迹、对于家庭的影响，以及社区服务的使用，如自助团体。例如，他们提及了病患的基本策略，不是以症状控制或维持生命为目标，而是以正常化为目标。在这个研究中，尽可能正常地过他们自己的生活（和过家庭生活）的过程不仅依赖于医疗的处置，也依赖于“病症的侵扰程度、养生法、对于他人患有同样疾病及疾病的潜在致命性的知识”（Strauss et al., 1984, p. 79）。如果这些疾病因素变得有侵入性，人们不得不“非常努力地为自己创造与一些正常生活的相似之处”（p. 79）。斯特劳斯等人的描述，即术语“生活的异常化与正常化”，是定性驱动理论丰富性和复杂性的一个卓越的示例，而这只有当考察者在资料搜集和分析期间运用理论敏感性的原则才能获得。

——Janice M. Morse

（王玥译　赵锋校）

参考文献

Glaser, B. G. (1978). *Theoretical sensitivity*. Mill Valley, CA: Sociology Press.

Glaser, B. G. (2002). Grounded theory and gender relevance. *Health Care for Women International*, *23*, 786-793.

Morse, J. M. (1997). Considering theory derived from qualitative research. In J. Morse (Ed.), *Completing a qualitative project: Details and dialogue* (pp. 2163-2188). Thousand Oaks, CA: Sage.

Strauss, A. L., Corbin, J., Fagerhaugh, S., Glaser, B., Maines, D., Suczek, B., & Weiner, C. (1984). *Chronic illness and the quality of life*. St. Louis, MO: Mosby.

理论（Theory）

在社会科学领域，几乎没有哪个术语比“理论”一词有更多样的且常常模糊的意义。这里至少涉及三组对照实例：与实践，与事实，与证据。相对于“实践”，“理论”常指被阐明的原则，而不是实际所做的。在此，存在着一项重要的含义且可以有两种解释。理论可以被当作一个令参与者相形见绌的典范，或者类似的事物，而如果人们仿效它，将能改变实践。另一种解释是理论可能是非现实的或仅仅理性的产物，而不被接受。更加调和的解释是：理论由规范性的原则组成，而这些原则给予实践以不带具体指

示的一般的指导。

第二组与事实的对照也可以有不同方向的价值倾向。理论可能因为是可疑的且是某种被避免的事物而被拒绝接受。这是一些经验主义者的态度。相反,"事实"可以指呈现了普通性的且具有系统误导作用的,由历史构建起的表象。基于此,理论化的角色在于穿过事实的表面,以识别出潜在的生成结构,并由此显现出根本的社会变迁是可能的。理论的这一含义主要来自马克思主义(Marxism)和批判理论【Critical Theory】。

理论的第三种意义涉及与经验或资料的对照。此处,理论是事实性的,既不含有价值判断的意义,也不含有思索的意义。尽管如此,其意义也有相当的变化。在谱序的一端是理论框架或方法,例子包括行为主义【Behaviorism】、功能主义、理性选择理论和符号互动论【Symbolic Interactionism】。另一端是关于特定类型现象及其发生条件的几组专门的解释原则。一个例子是教育社会学中的差化—极化(differentiation-polarization)理论。这个理论宣称,学生的分化——例如,通过班级编排、成绩分组、学生组合或学生选派——产生出他们对学校态度上的极化,即那些排序高的学生变得更积极,而那些排在底部的学生则更消极。当然,其他的理论处在两个极端之间。一个可能的例子是越轨社会学中的标签理论。具体地说,该理论认为,与一般的信念相反,给违规者贴标签和惩罚他们可能增加而非减少他们未来再次做违规的事的机会。然而,标签理论也可以作理论框架来使用,比如,有一种就强调研究贴标签的人同研究被贴标签的人同样重要,同时,要求暂时搁置关于对错的传统观点。

有时,一些修饰词会附着于理论。它们可能用以修饰一种特定的理论方法,如"批判的"或"理性选择"理论。其他的修饰词则标示出所涉及的抽象水平的对比,如"宏大"和"中层"理论的对比(Merton, 1968, p.1)。还有些修饰词是方法论的,指理论是如何产生的(如扎根理论【Grounded Theory】)。

——Martyn Hammersley

(王玥译　赵锋校)

参考文献

Hammersley, M. (1995). Theory and evidence in qualitative research. *Quality and Quantity*, *29*, 55-66.

Merton, R. K. (1968). *Social theory and social structure* (3rd ed.). New York: Free Press.

浓描(Thick Description)

人类学家克利福德·格尔茨使得浓描这一借自哲学家罗伯特·赖尔的术语,成为社会科学家们的通行术语(Geertz, 1973; Ryle, 1971)。两位作者都拒绝"淡"描,或那些试图将人类行动还原成可观察的行为的描述。浓描的需要来自于下述事实,即行动由文化意义所界定,而非其物理特征。例如,可以将神经抽动与眨眼相对照:表面的行为可能都一样,但是,第二个是一项行动,而第一个则不是。而且,一个眨眼可能既不同于对某一眨眼的模仿,也不同于某人反复进行的眨眼。如这些例子清楚地表明行动是有意义的,而意义依赖于背景——依赖于谁在做什么、何时、怎样、为何、对何种行动

或事件的回应、带来何种后果等。因此，格尔茨论说，合理的描述必须包括蕴含于所描述行动中的意义网。

他不仅将浓描与人类行动的行为主义【Behavioristic】的记录相对照，使其对照于将意义从其地域背景中抽象出来的结构主义者【Structuralist】的说明。他在人类学著作中写入了下面的话，“理论陈述附着于它们所支配的阐释之上，离开了阐释，理论既没有意义，也会变得乏味。”（Geertz，1973，p. 25）。浓描的任务既要去捕捉特定事件的复杂性，也要去指出它们更一般的文化意义。也许，浓描的最著名的例子是格尔茨的文章“奥戏：关于巴厘岛斗鸡的说明（Deep Play：Notes on a Balinese Cockfight）”（Geertz，1973）。在这篇文章中，他用对一件目击事件的细致分析，解释了巴厘岛文化的特征。

格尔茨提出，生产浓描应该是人类学研究的首要目标。他坚持认为，这样的阐释人类学不比物理或化学更少具科学性，但它确实是一种不同类型的科学。同时，他认为人类学家的任务类似于文学批评。据此，他建议，民族志【Ethonography】类似于尝试构造一个手稿的一种阅读，而这种阅读是“异文化的、缺乏原本的活力，同时，充满了省略、不连贯、可疑的校正和有偏见的评论”（Geertz，1973，p. 10）。

浓描这一术语已经为许多定性研究者所采纳，将其作为一种标示出他们作品的特征倾向的方式。同时，浓描还被用于说明定性研究的发现如何能够被概化。因此，古巴和林肯论说，定性研究的读者们也需要用浓描来评估研究发现对于其他背景的可转移性（transferability）（Guba & Lincoln，1985；也可参见 Gomm，Hammersley，& Foster，2000）。

——Martyn Hammersley

（王玥译　赵锋校）

参考文献

Geertz，C.（1973）. *The interpretation of cultures.* New York：Basic Books.

Gomm，R.，Hammersley，M.，& Foster，P.（2000）. Case study and generalisation. In R. Gomm，M. Hammersley，& P. Foster（Eds.），*Case study method*：*Key issues*，*key texts.* London：Sage.

Guba，E. G.，& Lincoln，Y. S.（1985）. *Naturalistic inquiry.* Beverly Hills，CA：Sage.

Ryle，G.（1971）. Thinking and reflecting. In G. Ryle，*Collected papers*，*Volume* 2. London：Hutchinson.

三阶（Third-Order）

参见阶【Order】。

阈效应（Threshold Effect）

若两个变量间的关系在达到其中一个变量的某个特定值或两个变量的特定值时发生重要的改变，阈效应即出现。分析者如想正确地描述两个变量间的关系，建立阈效应模型极为重要。

阈效应的一个示例是一个国家的经济

系统与该国的民主表现水平间的关系。假定人们可以测量一个国家的资本主义程度。这个测量可以是定序水平的，如由弗雷泽研究所之类机构提出的“经济自由(economic freedom)”指数。它也可以是定距测量，如政府开支占国民生产总值的百分比。以资本主义作为自变量【Independent Variable】，进一步假定民主作为因变量【Dependent Variable】可以用一个 0~100 的量表予以测量，其中，低分数表明该国较不民主(例如，缺乏对公民权和自由的尊重)，而高分数显示该国是一个比较民主的国家(例如，更可能推行自由和公平的选举)。那些相信在资本主义和民主的关系间存在阈效应的人，在理论上会认为资本主义仅在某个点以内有其作用，越过了这个点，一国越具有资本主义性质，它就越不民主。这里的理论是，高度发展的资本主义国家，即政府对经济的干预有较低的水平，将不能提供社会服务、社会安全网的福利开支或容许在该国进行全面的民主表达的法律执行。

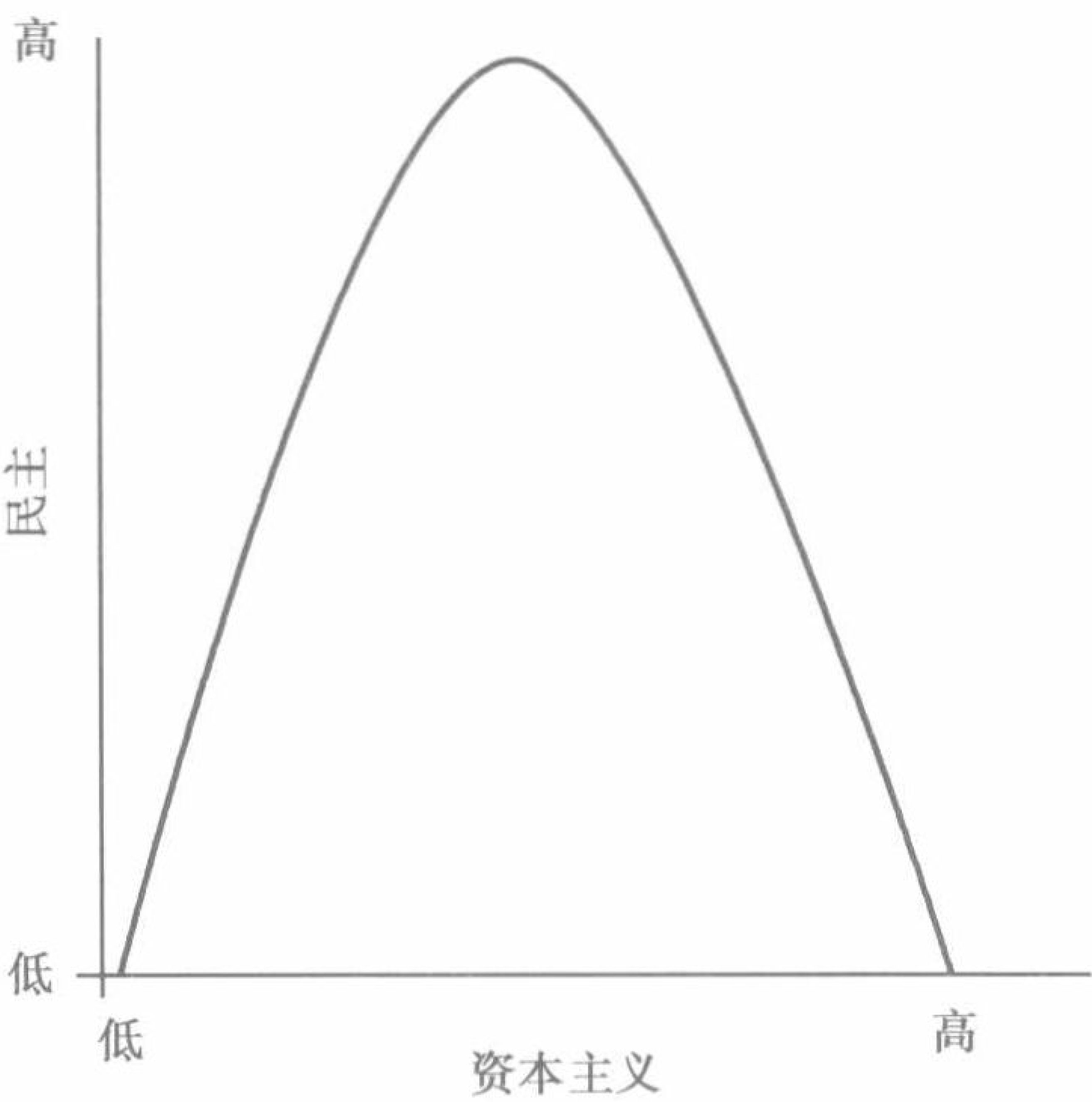


图 1　资本主义对民主的抛物线关系

使用最小二乘回归模型技术的经验研究显示，确实存在着一个统计显著【Statistical Significance】的资本主义对民主的阈效应，即超过这个阈效应，资本主义就会对一个国家的民主表现产生负面影响。图 1 显示，阈效应是一个倒 U 形曲线，即在一个较低的范围内随着资本主义水平的增长，经济自由促进民主表现直至某一点。越过了该点，关系出现反转，图中的正关系变为负关系，即更高水平的资本主义实际会降低民主表现，并形成倒 U 形曲线右侧的下行部分。

建模这个阈效应的正确方式是取自变量的平方——在此例中是资本主义——再将其作为第二个自变量加入模型。如果存在阈效应，那么线性项和平方项都会获得传统的统计显著水平 ($p< 0.05$，双侧检验)。计算发生阈效应时的自变量值——图中从正的斜率到负的斜率的斜率变化[称为“反转点(tipping point)”]——的数学公式可使用公式 $-b_1/2b_2$，其中，b_1 是自变量的线性项的最小二乘回归系数【Regression Coefficient】；b_2 是自变量平方项的最小二乘回归系数。

布林(Breen，1996，p. 50)给出了另一个阈效应的例子。仅当货币兑换率落在一个较高值和一个较低值的范围内时，它们才可以直接观察到；换句话说，有一个较高和较低的阈限。如果分析者试图解释货币兑换率，那么它同样是一个删截回归模型的例子。分析者可以使用 Tobit 估计量来估计删截回归模型。

无论例子怎样，如果分析者没有恰当地设定阈效应的模型，那么，他们在解释参数估计的结果时都可能误入歧途。

——Ross E. Burkhart

(王玥译　赵锋校)

参考文献

Breen, R. (1996). *Regression models: Censored, sample selected, or truncated data* (Sage University Paper Series on Quantitative Applications in the Social

Sciences, 07-111). Thousand Oaks, CA: Sage.
Brunk, G. G., Caldeira, G. A., & Lewis-Beck, M. S. (1987). Capitalism, socialism, and democracy: An empirical inquiry. *European Journal of Political Research*, *15*, 459-470.
Burkhart, R. E. (2000). Economic freedom and democracy: Post-cold war tests. *European Journal of Political Research*, *37*, 237-253.

瑟斯通度量(Thurstone Scaling)

瑟斯通度量是一个统称,涵盖了由心理计量学家路易斯·瑟斯通在1920年代研发的三项技术组成的系列。这些度量方法通常被称为:

1.等似间隔(equal appearing intervals)法。

2.连续间隔(successive intervals)法。

3.配对比较(paired comparisons)法。

在所有这些度量技术之下存在着一个关于态度测量的共同视角——比较判断律(the law of comparative judgment)(Thurstone, 1927)。

试想一下,你想比较多个物体的质量,但是没有现代的天平或者一套标准砝码。然而不必认为这样就不能精确地判断质量,人们可以估计物体的相对质量——从轻到重给物体质量排序。就每一个物体,一次比较它同其他物体的质量,再根据配对比较判断它们的相对质量,即可实现排序。在平衡横杆的两端各放一个物体,然后识别出较重的物体。一个完全系列的比较当可生成一个根据质量排序的物体集合。瑟斯通第一次有意识且系统地将同样的比较应用于心理对象,而不是物理对象。

一种旨在为心理学对象的值或强度赋予数值的度量方法必须反映出个体在判断多个刺激过程中做出的区分。潜在于比较判断度量法之下的要点是,相比于那些难于区分的对象,那些易于区分的对象应当被放置得更开(远)。

配对比较法包括,由所有受试者对对象的每一种组合作出明确的判断。随着对象数量的增加,这一任务变得十分艰巨,因为配对比较的总数呈几何增长。换言之,假定一组配对A和B等价于配对B和A,比较总数等于$N(N-1)/2$。4个对象会有6种配对方式,5个对象有10种,10个对象有45种。如果考虑顺序,比较总数的增长将更快:$N(N-1)$。4个对象将产生12组比较,5个对象需要20组,而10个对象需要90组。大量对象的排序会变得难以处理。

等似间隔法是一项在无须进行所有可配对的比较的前提下,捕捉个体对态度对象估计的技术。判别人被邀请来标定一组题项的刻度,而后再用校准后的量尺来研究总体。

该技术起始于一个数量较大的陈述集,而这组陈述反映了对某一对象的不同观点。将这些陈述呈现给一个判别人小组,请其中的每一个人将所有的陈述分别放入11个编号(从1~11)排序的堆(从“最赞成”到“中立”到“最不赞成”)中的一个,从而确定对对象赞成(或反对)的程度。这些初始的判断为最终量表的题项选择奠定了基础。

按照项目的平均赞同得分(中位数【Median】)和赞同得分的变差(四分位距【Interquartile Range】)对所有题项进行排序。舍弃判别人评定不一致的项目,最终的量表包括5~20个挑选出的题项,而这些题项不仅涵盖了整个赞同的范围,还以近似均匀的间隔被依次排列(依照判别人给出的中

位估计数)。

等似间隔量表趋向于排除极值点处的变差。而连续间隔法是作为等间隔度量法所做的假定(即实际上所有的间隔都是相对的假定)的纠正而得到发展的。连续间隔法为了捕捉极端的位置,允许存在不相等的间隔。

瑟斯通度量法在以下诸方面存在局限:

• 单维性假定。这些量表不能说明态度的多维度性质,反而将回答限定于单维结构。

• 劳动密集。在探索对于一个刺激对象的态度时,研究者首先必须创建一个巨大的,可以表现对对象的一系列意见的题项库。对那些移动题项,以构建最终量表的判别人而言,量表的创建涉及大量的工作。

• 判别人和受试者之间假定的相似性。如果从总体中选出的判断人有别于所要研究的受试者,这可能造成量表不能应用于研究总体。

由于瑟斯通方法潜在的弱点,今天这些度量技术更多成为历史研究的兴趣所在,而少有实践的应用。

——John P. McIver

(马妍译　赵锋校)

参考文献

Edwards, A. L. (1957). *Techniques of attitude construction*. New York: Appleton-Century-Crofts.

McIver, J. P., & Carmines, E. G. (1980). *Unidimensional scaling*. Beverly Hills, CA: Sage.

Scott, W. A. (1968). Attitude measurement. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology*. Reading, MA: Addison-Wesley.

Thurstone, L. L. (1925). A method of scaling psychological and educational tests. *Journal of Educational Psychology*, *16*, 433-451.

Thurstone, L. L. (1926). The scoring of individual performance. *Journal of Educational Psychology*, *17*, 446-457.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*,273-286.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, *23*, 529-554.

Thurstone, L. L. (1929). Fechner's Law and the Method of Equal-Appearing Intervals. *Journal of Experimental Psychology*, *12*, 214-224.

Thurstone, L. L. (1931). Measurement of social attitudes. *Journal of Abnormal and Social Psychology*, *26*, 249-269.

Thurstone, L. L., & Chave, E. J. (1929). *The measurement of attitudes*. Chicago: University of Chicago Press.

时间日记(Time Diary)

参见时间测量【Time Measurement】。

时间测量(Time Measurement)

时间是一社会产物。我们知道,所有的人类生活都与一定的节律联系在一起,而这些节律则来自某些多少具有规律性的“时间给予者”(它们既可能是自然的,也可能是人为的:天上的星座、太阳和月亮、教堂的钟、工厂的哨子、电视节目、家庭仪式)。不

同的时间给予者还化身于不同的权威结构、支配结构或互惠结构之中,而正是这些结构生成了每个个体在一天之中必须历经的多种混合的时间结构。因此,随着背景和环境的变化,每个人将以不同的方式体验到时间的历程。尽管如此,还是存在着一个大家公认的物理学时间,例如,用一个钟摆的摆动或一个铯晶体的周期震荡计数时间。不过,我们习惯于用日记【Diary】法,通过将活动与钟表时间相对照的方式来测量各种活动的持续。"时间预算调查"涉及样本【Sample】的日记,从而容许研究者据此估计一个给定总体在某个给定日历周期内的时间使用模式。超过 80 个国家已经收集了具有全国代表性的时间日记数据。绍洛伊(Szalai,1972)提供了一个涵盖 12 个国家的跨国日记数据;目前,多国时间使用研究(multinational time use study)为二手研究提供了 25 个国家的最近历史时期的微观数据(Gershuny, 2000)。

日记法可以反映日常活动的结构。日记簿是完整的和连续的,这样就使得所有试图掩饰某些活动的回答者必须刻意努力用其他的活动替代他想要掩饰的活动。日记容许考察活动发生的序列,或考察一天中活动发生的时间。日记上不仅可以记录下同时发生的活动,并且研究者还可以用它们来考察人们共同在场、相互依赖及共同合作的模式。然而,时间日记方法并非毫无问题。特别地,它会对于某些"社会不期许的、尴尬的和敏感的"行为予以低报,其中,包括个人卫生行为、性行为和洗浴行为(Robinson & Godbey, 1997, p.115)。

在时间日记工具的设计上存在着众多变化的样式(Harvey, 1999;就一般目的的设计给出了许多建议)。正如所有强行的(obtrusive)调查方法一样,在被访者因更大量和详细的日记而承受的负担(形成较好的数据但是较低的回答率)和通过所谓轻型日记工具实现的高回答率之间存在着冲突。(然而,无论哪种方法,其综括的统计结果似乎都相当稳定,Harvey, 1999)。特殊的问题源于个体活动中的日记变异。一个简短的(如一天的)日记会在人际变化方面给人一种伪造的印象,而实际只是个人内心的变化。一个较长的(如 7 天的)日记会透露更多的内容,但是也需要回答者付出更多。

日记的实施有两种可供选择的模式,分别是"昨日"法和"明日"法,"昨日"法是访谈员主导的过程,而"明日"法使用自填式工具。昨日法日记通常涵盖先前的几天,尽管记忆会变模糊,但在 2 天或 3 天的间隔之后,它仍可能收集到有用的信息;有时在访谈过程中,还需要克服由于经过了一个周末或短假带来的记忆困难。明日法的自填式日记是访谈员留给回答人自行记录的日记,同时,访谈员还指示回答人在整个日记周期中尽可能频繁地做记录,以取得最佳的记忆质量。这些明日法日记通常只包括 1 天或 2 天,但是荷兰和英国在 7 天日记方面已经具有相当成功的经验。有时研究会将由访谈员主导的昨日法和由回答人寄回自填式工具的明日法结合在一起使用。在这种情况下,明日法显示出较低的回答率,以及某种程度上有偏的回答。对于明日法日记,如果由访谈员来收回日记,那么回答会有实质性的改善。

进一步的设计问题涉及记录活动的时间框架。昨日法日记经常采用一种以事件为基础的记录("你什么时候起床?""你首先做什么?""你接下来做什么?")。对于自我记录,更合适的做法是使用一个日程表,而日程表以 5~15 分钟为一个时间间隔(不过,也曾使用过多达 30 分钟的时间间隔,而且间隔的长度可以随白天和夜晚的不同发生变化);回答人用直线、箭头或重复的符号

在日程表上标出每项活动的持续时间。

活动可以根据一个包罗的、"封闭的"预编码的活动清单予以报告,或者用一种"开放的"("自己的话")方式来报告。封闭的方法可能会过度限制回答人,活动类别的选项还会在一定程度上影响报告的水平。但是开放式会产生一个对应的问题:回答人会以不同的细节水平报告其活动内容,从而难以将回答人之间存在的真正的个体间差异分离出来。

在一天中的大部分时间内,回答人都会从事多项并行的活动。常规的做法是要回答人以一种主辅的方式记录活动:[先问,"你(那时)正在做什么?"再问,"在同一时间,你还做了其他事吗?"]有的研究则要求回答人以非主辅的方式记录所有活动,但回答人倾向于选择使用一种叙事的模式("首先我做了这个,然后我做了那个……"),而叙事的模式自然地包括了主辅的方式。这种主辅结构还能简化过程的分析。虽然主要活动是最常见的分析焦点,但是,不分主辅地将所有活动合并在一起呈现也是可以的,这依然可以覆盖全天的 1 440 分钟。

其他辅助信息包括在日记表的行或列中包含共同在场("谁和你一起参加了这项活动?")和当前位置(或移动模式)的信息。专门的日记研究还包括与每个事件或每个活动联系在一起的解释性信息或情感信息(如"这有用吗?"或"你有多么喜欢这个?")。

除了日记工具的设计,还存在着各种具体的抽样问题。季节性影响可能很重要;时间使用的代表性样本需要对一年有一个完整的覆盖。传统的明日法研究还包括了年龄低至 10 岁的孩子。

测量时间使用的一个可供选择的方法包括将"概略的估计"("你在……上花了多少时间?")作为问卷调查的一部分。这种方法不但是收集关于人们如何使用时间的信息最简单和最廉价的方法,在其他方面还有着广泛的使用(如在全国劳动力调查中)。但这种方法也有重大缺陷。针对由这种方法作出的有偿工作时间的估计的研究表明,估计存在着系统偏差,特别是高估了有偿工作时间(Robinson & Godbey, 1997, pp. 58-59)。

民族志【Ethnography】的研究经常涉及由研究者直接观察人群,并记录他们的活动。而当前,新的技术已经拓展了直接观察的作用,包括使用个体定位监视系统,以及闭路式或因特网摄像系统作为非强行的观察手段。观察方法有其优势,但是它们既昂贵又要求大量劳动投入,同时它们还可能危害个人隐私。此外,当人们对持续的监控有所察觉时,他们就会主动审查或扭曲自己的正常行为模式。

——Jonathan Gershuny

(马妍译　赵锋校)

参考文献

Gershuny, J. (2000). *Changing times: Work and leisure in postindustrial societies.* Oxford, UK: Oxford University Press.

Harvey, A. S. (1999). Guidelines. In W. E. Pentland, A. S. Harvey, M. P. Lawton, & M. A. McColl (Eds.), *Time use research in the social sciences* (pp. 19-45). New York: Kluwer.

Robinson, J. P., & Godbey, G. (1997). *Time for life.* University Park: Pennsylvania State University Press.

Szalai, A. (Ed.). (1972). *The use of time.* The Hague: Mouton.

时间序列截面(TSCS)模型(Time-Series Cross-Section [TSCS] Models)

时间序列截面模型(TSCS 模型)的目标在于查明时间序列截面数据(TSCS 数据)(常称为定群【Panel】数据)潜在的因果过程。人们可以以三维"数据立方体(data cube)"的形式呈现此类数据。在此立方体中,个案、变量和时点构成了三个不同的维度(参见图 1)。相比于(a)横截面数据【Cross-Sectional Data】组和(b)时间序列【Time-Series】数据组,TSCS 数据组含有更多信息,横截面数据组由单一时刻各个个案的变量的信息构成,而时间序列数据组由单一个案在不同时刻上的变量的信息构成(例如,图 1 立方体最上一层)。

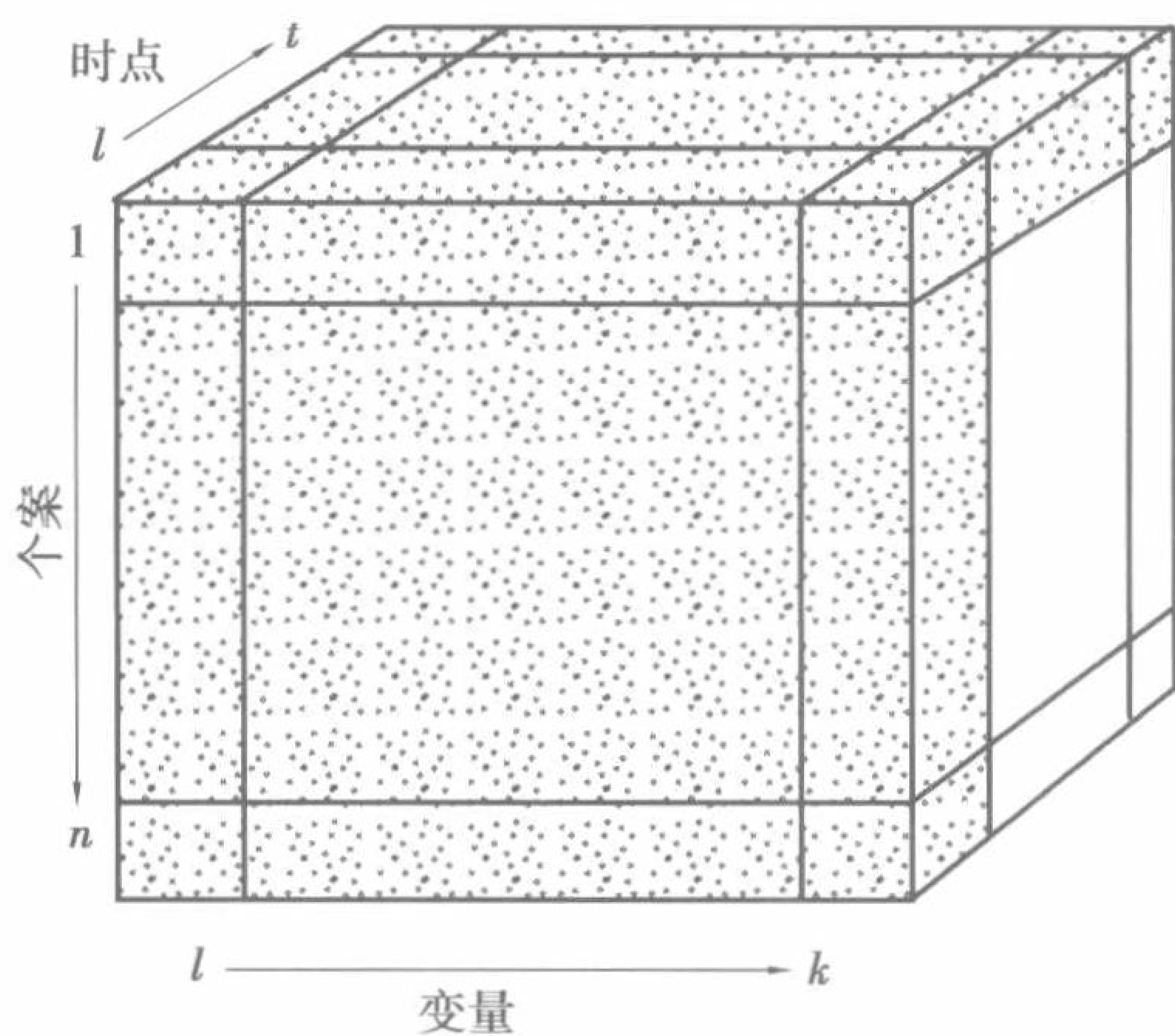


图 1 呈现一组时间序列截面数据的数据立方体

研究者有时不得不因为目标群体的个案数量少而收集 TSCS 数据组。例如,任何关于欧洲国家的截面研究都受这样的事实制约,即它的个案数少于 30,这就相当严重地限制了人们运用多元统计分析的能力。通过在 7 个不同时点而非一个时点上收集这些个案数据,研究者可能获取充分的额外观测值以便更好地统计分析。

TSCS 数据组也因为其分析能力而受到关注。例如,使用 TSCS 数据,人们可以估计一自变量是否随时间的变动影响因变量(对横截面数据不可能有此分析),或者是否一自变量随个案的不同,效应也不同(对单一个案时间序列数据不可能有此分析)。此外,研究者还常常使用 TSCS 数据以便控制基于个案的未测量到的变量(case-stable unmeasured variables),在多元分析中,此类未测量到的变量是未测量到的异质性中的重要类型。

然而,TSCS 数据提供的分析便利并非无限制。要从时间序列和横截面数据的汇聚中得益,其前提需假设某种自变量效应(包括非线性和非累加的效应)的不变性。这些效应需要在不同的样本上,不同的时点上或者两者同时保持恒定(实践中,多数研究者同时假定两者不变)。此外,在 TSCS 分析中,异方差性【Heteroskedasticity】和自相关【Autocorrelation】都是统计推论【Statistical Inference】的潜在威胁,前者主要被视作截面统计分析的潜在问题,而后者主要被视作时间序列统计推论的主要问题。

大 n,小 t 的 TSCS 模型

由于 TSCS 数据提供了多种分析的可能,于是有许多可用的 TSCS 模型。然而,有两个一般类型的模型,分别对应于个案(n)的数量是否实际超过时点(t)的数量,或者相反。大 n 定群研究,如美国国家青年的纵贯研究和美国家庭收入动态变化的定群研究,都有成千上万的个案,而时点相对较少(如<15)。使用这类数据的研究一般视个案中未测量到的异质性为估计偏倚和估计不一致的最主要的潜在来源,并运用固定效应[Fixed Effcts, FE]或随机效应[Random-Effects, RE]解决这个问题。两组估计方法

都来源于分析模型：

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \cdots + \beta_k X_{kit} + \alpha_i + \varepsilon_{it}$$

其中，i 表示个案号；t 代表时点，数值下标表示测定的自变量（X）；β_0 代表方程的截距；ε_{it}代表均值为零，方差【Variance】恒定，同方程中的自变量无关联【Association】的干扰项【Disturbance Term】；α_i代表未测量到的变量的效应，该效应就每个个案而言不随时点变化，但是随个案的不同而变化。例如，在个人是个案的研究中，这项代表稳定的但未测量到的变量，如一个人的天生能力或外表的吸引力，并影响个案在因变量上的取值。

模型的 FE 估计项中包含有对处于方程自变量的 α_i 的估计。分析者既可以通过作出 n-1 个虚拟变量表示 n 个个案，直接估计 α_i，也可以通过分析被测变量在每个个案上的“时间减损”效应，间接估计 α_i（Wooldridge, 2000, pp. 441-446）。因为 α_i 代表每个个案随时间稳定的变量，所以使用固定效应估计项使得估计稳定的，已测的自变量的偏效应成为不可能，例如性别和人种/民族。不过，多数研究者同意这是为了控制众多未测量到的变量的效应而必须付出的微弱代价。［尽管一阶差分估计项（参见 Wooldridge, 2000, pp. 419-425）同固定效应估计项很接近，却不经常使用，即使在某些情况下，它可能有更好的统计属性；Wooldridge, 2000, pp. 447-448.］

对大 n，小 t 的 TSCS 数据组，RE 估计项假定 α 为随机变量，同时将其置于方程的误差项，因此产生了一混合的误差项 $u_{it} = \alpha_i + \varepsilon_{it}$。这一方法有三项重要的隐含意义。第一，除非在分析中由 α 代表的变量同测得的自变量不相关，否则 RE 估计项就是有偏的且不一致的。许多研究者发现难以假定 α 满足此项条件，进而偏向 FE 模型。例如，外在吸引力（通常未测量）就很可能同个人收入（通常已测）正相关。第二，当基于个案的变量被置于混合误差项，人们必须使用可行的广义最小二乘技术（feasible generalized least squares estimation techniques, FGLS）将下列事实纳入考量，即就每个个案而言干扰项可能在时间维上相关。幸运的是，许多统计软件（如 SAS, Stata）包含有执行此类 FGLS 方法的程序。第三，在混合误差项中包含 α 使得估计随个案变化的已测自变量的偏效应成为可能，当时还不能估计随时间变化的偏效应，因为在使用 FE 模型时，这是不可能的。

如果在分析中 α 同已测得的变量不相关，那么 RE 模型比 FE 模型会更有效，因为除了标准的普通最小乘方（OLS）估计之外，它仅需估值一个额外的量（α 的方差）。相反，当 FE 要估计每个 α_i 的值时，就需要估计额外的 n-1 个量。研究者常用 Hausman 检验来确定 α 是否同分析中已测得的变量不相关（Wooldridge, 2002, pp. 288-291），进而确定使用可能更有效的 RE 估计项是否合理。类似于未测量的个案效应，上面的模型还可以扩展以包含一个未测量的时间效应项。这时，模型写为：

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \cdots + \beta_k X_{kit} + \alpha_i + \nu_t + \varepsilon_{it}$$

其中，ν_t 代表在给定时点上随个案不同而恒定的未测量变量的效应。服从上述仅包含 α_i 项模型相同的优点和缺点，人们可以使用 FE，RE 或合并 FE 和 RE 来估计此模型。

小 n，大 t 的 TSCS 模型

研究者分析带有少量个案和众多时点的 TSCS 数据组，一般把异方差且自回归的干扰项视作有效推论的主要威胁，同时利用

每个个案的众多观测值来估计 FGLS 模型，进而调整这些威胁。最少，这类模型合并了以个案为单位的异方差性和不同个案共有的自相关模式，但是还有其他许多可能的模型。例如，一个流行的替代模型 Parks-Kmenta 模型，合并了以个案为单位的异方差性，基于具体个案 AR1 水平上的自相关，以及针对不同个案，干扰项在具体时点间的相关。复杂数据结构的估计要求在众多时点上有数据，然而，蒙特卡罗试验说明当时点的数量小于 40 时（Beck & Katz，1995），Parks-Kmenta 模型会产生严重的有误导的统计检验。大多数统计软件包含有一组用以分析小 n，大 t 的 TSCS 数据的备选模型。

——Lowell L. Hargens

（赵锋译校）

参考文献

Beck, N., & Katz, J. N. (1995). What to do (and not to do) with time-series cross-section data. *American Political Science Review*, *89*, 634-647.

Kmenta, J. (1986). *Elements of econometrics* (2nd ed.). New York: Macmillan.

Wooldridge, J. M. (2000). *Introductory econometrics: A modern approach*. Cincinnati, OH: South-Western.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge: MIT Press.

时间序列数据（分析/设计）（Time-Series Data [Analysis/Design]）

时间序列数据常出现在政治经济学、宏观经济学、宏观政治社会学等相关领域的研究中。时间序列数据是对某个实体（如一个国家），在规则的间隔期间重复地观测。它们不同于横截面数据【Cross-Sectional Data】，后者由对一组实体（可能是个人或国家）样本在一个时间点上的观测构成。尽管有些分析者分析单一时间序列[参见 Box-Jenkins 建模法（Box-Jenkins Modeling）]，大多数研究者则运用时间序列数据研究一个因变量同一个或多个自变量间的关系，类似于典型的回归【Regression】。简单起见，这篇介绍仅考虑单个自变量，不过，这里所说的一切很容易推广到多元回归【Multiple Regression】的情形。因为时间序列模型是回归模型的一个类别，所以与回归模型相关的所有问题也都适用，包括检验、设定【Specification】等。本文首先通过讨论一些时间序列数据的例子指明时间序列数据分析的具体问题；最后指出模型设定和解释的一些问题[也可参见滞后结构【Lag Structure】]，并在结论部分扼要讨论时间序列数据分析的未来方向。

时间序列数据的例子

宏观经济学家是时间序列数据最常见的应用者。一般，他们拥有对多种经济现象的月度或季度观测，并且他们有兴趣将某些政策工具（如货币供给）同某些经济利益变量，如 GDP 的增长联系起来。政治经济学家和社会学家也从事相似的分析，他们的兴趣在于政治变量，如权力联盟的政治取向对经济产出的关系，或者经济变量，如失业率或通货膨胀，对政治后果的关系，如对总统的拥护或选举结果。时间序列数据也出现在许多其他领域，如冲突研究，它分析防御支出随时间变化的原因。在所有这些研究中，分析者需要清楚地了解手中数据的频次，每日的、月度的、季度的还是年度的；不

同的模型适用于较高或较低频次的数据。

本篇介绍仅限讨论固定式时间序列(stationary time series)。换言之,如果该时间序列的各种统计属性不随时间变化,它就是固定式的;尤其,固定式序列的最佳长期预测是它的总均值。非固定式序列涉及更艰深的数学,非固定性的原因常包括某种时间趋势。分析者需谨慎,他们要限制自己仅用固定式方法分析固定式时间序列数据。

尽管时间序列数据指对单一单位的观测,但它们可以被推广至时间序列截面数据【Time-Series Cross-Sectional Data】,其中,分析者研究多个单位的时间序列,通常有一组国家的年度观测数据。时间序列截面数据有许多时间序列的属性,因此许多同时间序列模型估计相关的问题也与之有关。定群【Panel】数据也与时间序列数据有关;此处,分析者通常仅有对大量个体的少量重复观测。尽管定群数据表现的估计问题同时间序列的问题有关,它的数据还是充分地不同,从而必须应用不同的模型和方法。

估计

固定式时间序列数据应用标准回归方法予以估计。如果数据满足所有此类模型的假定,那么普通最小二乘【Ordinary Least Squares】具有其通常的最佳属性。但是时间序列数据通常难以满足一项关键假定,即每项观测值相互独立(或者,等价地,每个观测值的误差项独立)。在较短的间隔期(如1个月)观测的时间序列数据一般高度相依;在相当大的间隔期(如10年)观测的数据可能相互独立。

相依为估计带来严重的后果。幸而,检验数据的相依和纠正相依的问题都很简单。但是,若不检验和纠正相依问题则会导致严重的无效估计和高度不精确的标准误。尽管违背某些假设不会有严重的后果,但是违背误差项独立的假设可以轻易地产生几百个百分点错误。因此,时间序列分析者必须非常谨慎地检验和纠正观测值间的相依。典型的假定是观测值同下一个时期的观测值间有最大关联,随着观测期的间距增大,两个观测值间的关系衰减。在某个时期的观测值的误差项记作 ε_t,假设简单回归模型:

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

关于误差项最简单的假设是它们服从一阶自回归(a first-order, autoregressive, AR1)模式,即

$$\varepsilon_t = \rho\varepsilon_{t-1} + \nu_t$$

其中,假定 ν_t 之间相互独立,ρ 表示误差项之间关联的程度。

分析者可通过估计回归模型并计算残差【Residual】来检验假设 $H_0: \rho = 0$,即误差项是独立的。然而,可以用有名的Durbin-Watson统计量【Durbin-Watson Statistic】来判定独立,或者更简单地,通过做OLS的残差同它们的滞后期,加上所有其他自变量的回归,再判定这个辅助回归的系数的显著予以检验。这是Lagrange乘数检验最简单的例证,此类检验在时间序列分析中非常有效。

如果独立的虚无假设没被拒绝,就可用OLS(假设其他的假定得到满足)估计时间序列数据。反之,如果独立的虚无假设被拒绝,似通常的情形,OLS估计将发生严重的缺陷。一个好的备选方法是广义最小二乘【Generalized Least Squares】。就带有一阶自回归误差项的时间序列数据而言,广义最小二乘实际上就是有名的Cochrane-Orcutt方法。这种方法有五个步骤:

步骤1:应用OLS于基础回归方程。

步骤2:应用从回归中得到的残差估计

ρ(通过做当期残差项对滞后残差项的回归)。

步骤3:将该估计值用于获得观测值间的“伪差值(pseudo-difference)”(此即,用当前的观测值减去估得的 ρ 乘以前期观测值)。

步骤4:以观测值间的“伪差值”为因变量,重做OLS,估计新的 ρ 值。

步骤5:重复步骤3和步骤4,直至 ρ 的估计值不再变化。

这一方法给出了 α 和 β 很好的估计,以及正确的标准误。

通常,我们将上一期 y 的观测值,滞后的因变量 y_{t-1} 加入设定,给出

$$y_t = \alpha + \beta x_t + \overline{\omega} y_{t-1} + \varepsilon_t$$

尽管Durbin-Watson检验对于滞后的因变量【Variable】不再有效,但是上面讨论过的Lagrange乘数检验依然是一个检验误差项是否独立的好方法。通常情况下,在滞后因变量模型中,误差项是独立的,这时,用OLS对模型有很好的估计。不过,分析者必须使用Lagrange乘数检验检查独立假设:如果误差项相依,OLS不再有一致的结果。

如果 x 和 y 不再固定,那么所有的标准方法都失效。简言之,分析者可通过做变量 x 对它的第一滞后项的回归,再检查回归系数的估值来估定固定性(对变量 y 亦然)。如果系数接近1(单位根【Unit Root】),那么分析者必须应用更加复杂的方法,远超本篇介绍的范围。然而,有一个模型通常对这类数据很有效,即误差校正模型【Error Correction Model】。该模型假设变量 x 和变量 y 间如同它们之间的短期关系一样,存在长期均衡的关系,由此,被估模型写为

$$\Delta y_t = \alpha + \beta \Delta x_t - \phi(y_{t-1} - \gamma x_{t-1}) + \varepsilon_t$$

方程圆括号中的量表示系统背离均衡的程度,而 ϕ 是系统调整回向均衡状态的速率。在许多条件下,此模型可以用OLS予以估计。然而,非固定式序列是复杂的,因此,分析者必须非常谨慎地使用合适的方法和模型于此类高技术领域。

最后,值得注意的是上述陈述限于这些模型,其中的效应或者都被间接地感受,或者作用于一期滞后项。当然,将其推广到更复杂的滞后结构【Lag Structure】是容易的。通常,一期滞后项可用于年度数据,不过,季度或月度数据常常要求更加复杂的滞后结构。这时检验的原则依然相似。还有许多额外的工具可用,如相关图,可以辅助分析者确定有多少滞后项包含在模型中。

模型设定和解释

在简单回归中,回归斜率 β 告诉我们,如果变量 x 变化一个单位,可以期望变量 y 变化 β 个单位。就简单时间序列模型而言,这是正确的,但是这类模型也可有更复杂的设定。例如,在滞后因变量模型中,变量 x 上1个单位的变化联系着变量 y 上 β 个单位初始变化,但是变量 x 变化的效应会持续。假设变量 x 在下一个时期回到初始值。尽管如此,在这个时期,因为滞后的 y 现在已高出 β 个单位,所以当前的 y 会高出 $\beta\omega$ 个单位。在更下一个时期,滞后的 y 会高出 $\beta\omega$ 个单位,因此,当期的 y 将高出 $\beta\omega^2$ 个单位,以此类推。该函数的曲线被称为脉冲响应函数。

作为选择,我们可以允许变量 x 变化1个单位,并永远保持在新水平上(水平变化分析)。接下来,变量 y 如何响应?最初,变量 y 随之变化 β 个单位。在下一个时期,滞后的 y 高出 β 个单位,但是变量 x 保持高出1个单位,所以当期的 y 高出 $\beta+\omega\beta$ 个单位。接下来的时期,变量 x 仍然高出1个单位,而滞后的 y 现在高出 $\beta+\omega\beta$ 个单位,因此,当

期的 y 现在高出 $\beta+\omega(\beta+\omega\beta)=\beta+\omega\beta+\omega^2\beta$ 个单位。这一几何级数如此持续下去，运用简单的算术可知，变量 x 上 1 个单位的变化最终关系到变量 y 上 $\beta/(1-\omega)$ 个单位的变化（假设 $\omega<1$，这是固定性的必要条件）。因此，如果 ω 较大（比如 0.9），那么变量 x 上 1 个单位变化长期影响 10 倍于其短期影响 β。必须将这一解释牢记。所有时间序列模型都一贯需要联系脉冲响应和水平变化分析给予解释（要谨记所有使用的系数都是估计得到的，因此估得的效应有置信区间）。

这种分析非常有益于考虑时间序列模型的恰当设定（不同于前面部分讨论的估计的计算障碍）。由此，尽管上面讨论过的广义最小乘方法可以正确地处理估计问题，脉冲响应分析却显示这一模型的潜在缺陷。尤其变量 x 的一单位变化对变量 y 有 β 个单位的直接影响，同时，此种影响即刻消失。但是，误差项 ε 上 1 个单位的变化对变量 y 有 1 个单位的直接影响，而这种影响在每个周期上以 $(1-\rho)$ 的速率衰减。因此，误差项有非常复杂的动力机制，感兴趣的变量 x 却被设定仅有直接影响，而不存在动力机制。尽管这可能是对过程的正确描述，表面上看起来却难以置信。另一方面，滞后因变量模型显示在变量 x 和误差项上存在相同的动力机制和衰减率；表面上这种设定更加合理。

显然有许多可能的模型设定需要分析者予以考虑。这些模型可以根据它们的理论属性以及它们拟合数据的程度而加以比较。无论哪一种比较，带滞后因变量的模型总是有力的竞争者。因此，重要的是记住包含滞后因变量和相依误差项的模型根本不能用 OLS 予以一致的估计（同样，任何起始于 OLS 估计的广义最小乘方法也将是错误的）。所幸，应用 Lagrange 乘数检验可以容易地检查在带滞后因变量的模型中误差项是否相关；恰好，此类检验的工作方式带有 OLS 残差。即便这类检验常会显示没有遗留的误差项相关，分析者在进行下一步工作前，也要严格地应用它们。

新方向

对单个固定式时间序列的分析目前已有很好的理解。通常 OLS 对此类序列可以发挥很好的作用，同时，严格的 Lagrange 乘数检验策略总是能指出何时 OLS 不能很好地适用。一旦 OLS 可以很好地发挥作用时，时间序列模型就容易估计，而且分析者可以专注于发现恰当的滞后结构（有时是理论地，通常则是经验地），以及通过脉冲响应和水平变化分析来研究被估模型的属性。理论兴趣依赖于待估的特定滞后结构的选择。

在过去 10 年间，许多研究注意非固定式时间序列，尤其是那些带有单位根的。经济学家致力于应用这些模型把时间序列分解成趋势和循环，而这些分解很久以来就吸引着宏观经济学家。研究者的注意力也聚集在分析多个时间序列上，例如多国之间的汇率，以便考察它们所显出的趋势和循环是否是共有的抑或有区别。所有这些模型的估计都存在技术困难，但是目前标准的时间序列软件都可以实现相应的估计。就那些试图解释单个非固定的因变量的分析者而言，简单误差校正模型通常足以满足要求（而且相应的检验可以显出它是否满足要求）。误差校正模型的优势在于它可以由 OLS 估计。

最近，许多注意力聚集于分析定群数据和时间截面数据上。定群数据通常对同一个体具有相对少量的重复观测，因此，应用成熟的时间序列模型于此类数据常常行不通。但是，研究者必须将注意力引向同一个体随时间变化的观测值间的相依。而时间

序列截面模型通常在每个单位上有重复的观测值,进而可以应用某些时间序列方法和模型来分析数据。对有志于比较方法的分析者而言,无论是经济学、社会学还是政治科学的,这都是一个极富吸引力的领域。最近,分析带有两分或其他限制的因变量的时间序列截面数据形成了一股风潮;这股风潮主要发生在政治冲突的研究中。以上所述都是热点研究的领域,但此类研究的大多数都始于单一时间序列分析。

——Nathaniel Beck

(赵锋译校)

参考文献

Beck, N., & Katz, J. N. (1995). What to do (and not to do) with time-series cross-section data. *American Political Science Review*, *89*, 634-647.

Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.

Harvey, A. (1990). *The econometric analysis of time series* (2nd ed.). Cambridge: MIT Press.

Mills, T. C. (1990). *Time series techniques for economists*. Cambridge, UK: Cambridge University Press.

Tobit 分析(Tobit Analysis)

在定义 Tobit 时,詹姆斯·托宾(Tobin, 1958)应用了一段关于无可能少于无的文字。Tobit 容许分析来自删截数据【Censored Data】,从而带有以零为最低值的自变量【Dependent Variable】,同时避免违背普通最小二乘【Ordinary Least Squares, OLS】关于自变量连续和非有界的假设。就此类自变量,OLS 可能产生负的预测值,并预测出不存在的值。应用 OLS 的其他后果包括有偏的系数和错误地得出统计上不显著的系数。获得负的预测值是 Tobit 解决的最普遍问题,此外,Tobit 可以分析带上限、上限和下限及中段界限的自变量。

Tobit 结合了 Probit 分析【Probit Analysis】和 OLS。它的系数容许辨识对某个因变量的两种效应:(a)对于超过其极限值的概率的效应;(b)对于超出极限的变化值的效应。例如,邻区的犯罪率不可能小于零,但是它们可以有任意正值。少量地区可能有零犯罪率。OLS 会预测出荒谬的负的犯罪率。操控 Tobit 系数容许辨识自变量对那些犯罪率为零的邻区有犯罪发生的概率的影响,以及自变量对有非零犯罪率地区的犯罪变化的影响。就那些超出某个限度,而不能予以测量的因变量取值,Tobit 对之作出调整,同时给出自变量效应的无偏估计。

Tobit 是大样本技术。个案对变量的比率应当大于等于 10。该技术对自变量的测量敏感。自变量应有相似的范围,即使变量的转换是必需的。它对异方差性的敏感可以由程序(如 LIMDEP)予以检验和调整。尽管有时 Tobit 模型比泊松回归【Poisson Regression】能更精确地模型计数变量的分布,但它要求因变量在极值外保持连续。Tobit 要求给予它所给出的效应以相同的符号。在一个无犯罪地区内发生一项犯罪的概率的期望变化必须相同于在有犯罪地区内发生犯罪的期望变化。否则,就要应用不同的技术。

回归截距与 Tobit 系数同它们的自变量之积的和可能为负。Tobit 系数是对观测到

的("潜在的")变量的效应。它们的分解容许对因变量的观测值的效应作出解释。系数的大小则依赖于自变量的测量尺度。更大的系数并意味着相应的自变量更重要。修正龙斯克(Roneck,1992)的伪标准化系数,即通过Tobit系数乘以相应自变量的标准误,可以对测量尺度作出调整。用每个系数除以因变量的标准误或Tobit模型的标准误都是无效的调整。

——Dennis W. Roneck

(赵锋译校)

参考文献

McDonald, J. F., & Moffitt, R. A. (1980). The uses of Tobit analysis. *Review of Economics and Statistics*, *62*, 318-321.

Roneck, D. W. (1992). Learning more from Tobit coefficients. *American Sociological Review*, *57*, 503-507.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, *26*, 24-36.

容差(Tolerance)

容差是一个帮助判定某个回归【Regression】模型中多重共线性【Multicollinearity】程度的量。它表示某个自变量【Independent Variable】的变化量统计上不能由模型中其他自变量的变化予以解释的程度。如果容差=1.0,那么问题中的自变量同模型中的其他自变量完全线性独立。如果容差=0.0,问题中的自变量则同模型中的自变量完全线性相依【Linear Dependence】。容差值越高,多重共线性则越不可能存在。在许多统计软件中,容差都是常规的计算内容。通常,分析者可以做自变量j对模型中所有其他非j的自变量的回归,获得能说明X_j变化的R方【R-Squared】。然后,根据这个辅助回归,再计算$R^2_{X_j}=1-$容差。例如,如果$R^2_{X_j}=0.65$,那么容差=0.35。

——Michael S. Lewis-Beck

(赵锋译校)

*也可参见方差-膨胀因子【Variance-Inflation Factors】。

总体调查设计(Total Survey Design)

总体调查设计是能够影响一项社会调查的估计结果的所有要素:调查中备选的人群样本【Sample】,那些实际被收集了数据的人的特征,问题的设计,以及用于提出问题和得到答案的程序。

总体调查设计的概念适用于设计目的在于产生某些总体【Population】的统计描述的社会调查。这类调查通常具有如下特点:(a)有一个被描述的总体的样本,而非整个总体,被选中以给予信息;(b)实际上,仅有部分被请求给予信息的人这样做,他们就是回答者【Respondent】;(c)数据由被选回答者给出被问问题的答案构成;以及(d)问题和回答(question-and-answer)的过程或者通过一个访谈员问问题,再记录答案,或者通过让回答者自己读纸面或计算机屏显的问题,再将答案记录在纸面上或直接记入计算机中的方式来完成。社会调查诸要素中的

每一项都能潜在地影响调查结果描述被研总体的完善程度。例如：

a.谁实际上有机会被选入样本，它们是怎样被选中的对于样本反映其所取自的总体的可能性特别重要。

b.那些实际上回答问题的人可能会系统地以同那些不能作答或选择作答的人不同的方式影响调查结果（见无回答【Nonresponse】）。

c.样本量，实际参与回答的人数将影响由样本结果做出的估计精度。

d.哪些问题被问，它们设计的完善程度都将对问题答案提供有效信息的程度，即调查所欲测量的内容，具有重要影响。

e.如果访谈员也是数据收集要素之一，那么访谈【Interviewing】的质量——他们多大程度上发挥了他们的作用——会影响最终数据的精度。

f.如果问题是自填式的，由回答者自己阅读问题，并给出答案，那么问题的形式和顺序设计的好坏会影响调查数据的质量。

总体调查设计的基本假设是，来自一项抽样调查的数据由所用方法中设计最差的项目决定。设计抽样调查时，实施调查时，或是评估调查数据时，重要的是检视所有能够影响数据质量的要素，即总体调查设计。

——Floyd J. Fowler, Jr.

（王玥译　赵锋校）

参考文献

Fowler, F. J. (2002). *Survey research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Groves, R. M. (1989). *Survey errors and survey costs.* New York: Wiley.

录音转录、转录（Transcription, Transcript）

尽管现在可以用数字形式储存、编码和检索音频和视听（audiovisual）资料，然而大部分以访谈为基础的定性研究【Qualitative Research】仍然依赖于录音资料到文本形式的转录。因此，准确的、高质量的、逐字逐句的转录是建立定性研究可靠性和可信性（严格的）所必需的。最大的问题在于此种错误，即颠倒和严重地改变录音所说的意义，因为它们可能导致研究者产生错误的诠释和错误的引用访谈回答。一篇对转录错误来源的评论文章（Poland, 1995）提出，最常见的转录错误来自改变措辞，即使是最表面的改变，仍可能产生原意的改变。这些错误可能发生于，例如，某些不同意义的词被他人误会有同样的发音，［例如，“咨询（consultation）”和“对抗（confrontation）”］；在模糊的上下文中插入标点（例如，句子“I hate it, you know. I do.”与句子“I hate it. You know I do.”）；或者人们模仿其他人的说话（或他们在脑海中的对话），因此，观点并不真的像它们表面显示的那样。转录者的疲劳、低质量的录音（由于背景嘈杂、使用低质量的录音设备和手机、话筒放置的位置不好）以及难懂的口音和文化上特定短语表达，都可能产生转录错误。当录音资料由专业的转录员转录而非开展访谈的研究者做时，由于缺乏对主题的熟悉，或对访谈回答者及访谈过程本身的熟悉常会造成另外的问题。

增强转录质量的措施，包括用高质量录音录制；注意转录者的选择和培训；使用记号系统来标识犹豫、强调以及口头互动的其他方面；成员确认【Member Validation】；以及对转录质量的后续检查程序。原声转录所需要的时间和精力（实际是关于原声的程度的决定）依赖于研究者的要求。例如，那

些进行谈话分析【Conversation Analysis】者的需要通常超过那些兴趣只在录音内容的主旨的定性研究者,因为前者要分析自然发生的谈话的结构和形式。

对转录质量问题的关注不能走向极端,因为研究者应当理解在一种介质到另一种介质的转换过程中,必定会产生某些损失。纸上的文本不可能再造出访谈的整个要旨和感觉,而那些根据访谈录音带来衡量转录准确性的人,可能清楚地记得大部分非言语(和一些言语)的交流已经从访谈录音中消失了。确实,在关于转录的很少但不断增加的文献中,一些声音已经开始认为,对"原声"转录的共同强调存在问题。他们主张更加反思的立场,强调把转录视为解释行为(包括资料的选择性强调和编码)的重要性,而不仅仅是一个最大化准确性的技术问题。

最后,由于口头语言和书面语言间在交流能力的标准上存在着的固有差异,坚持原声转录,即在出版的研究中使用引号,还可能不经意间将回答者表现成缺乏清晰表达能力的人,而如果是回答者自己写出相关主题,则不会这样。特别是在跨文化和阶级间研究中,这点有着重要的政治和伦理后果。

——Blake D. Poland

(王玥译　赵锋校)

参考文献

Lapadat, J. C., & Lindsay, A. C. (1999). Transcription in research and practice: From standardization of technique to interpretive positionings. *Qualitative Inquiry*, *5*(1), 64-86.

Mishler, E. G. (1991). Representing discourse: The rhetoric of transcription. *Journal of Narrative and Life History*, *1*(4), 255-280.

Poland, B. D. (1995). Transcript quality as an aspect of rigor in qualitative research. *Qualitative Inquiry*, *1*(3), 290-310.

Poland, B. (2001). Transcription quality. In J. Gubrium & J. Holstein (Eds.), *Handbook of interview research*. Thousand Oaks, CA: Sage.

转换函数(Transfer Function)

参见 Box-Jenkins 模型【Box-Jenkins Modeling】。

变换(Transformations)

变换常用于改变变量【Variable】的特征,如通过减少变量的类别数,合并两个或多个变量,或者改变变量的公制。在执行实际的数据【Data】分析之前,相当数量的时间要花在数据处理过程中。数据处理的主要过程就涉及变量的变换。

变换常用于重新归组定类或定序数据,以获得更少量、更有用的类别数。如果单变量统计学显示只有极少观测值落入变量的某些类别上,那么有理由将它们同相似的类别合并起来。就定类数据而言,任意类别都可以合并在一起,但是,对定序数据,只有毗邻类别才可以归并。重新归组的决策要在两种情形间达成妥协:一是变量有多个类别,其中某些类别只有极少观测值;另有者,请将数据并入较少的类别,从而导致有用信息的损失。

变换也常用于改变变量的编码

【Coding】,而无须重新归组变量。例如,如果两分类别被转换作 0 和 1 的编码,对数线性分析会更易于解释。带有三类或更多类别的定类变量常常被转换成一系列仅有两种结果(0 或 1)的虚拟变量【Dummy Variable】,以便用于多元分析【Multivariate Analysis】(Afifi & Clark, 1999)。此外,变换也常用于改变定距或定比数据的单位,加总或合并变量以生成新的尺度或者其他的总和分,生成几个变量的合并,以及综括跳跃模式。

最简单的运算之一是算术变换。所有用于数据处理和分析的软件都可以执行加(+)、减(-)、乘(*)、除(/)以及指数(**)(变量的乘幂,如取平方)运算。除了算术运算,大多数软件包还有众多的函数,可用于数据的筛选、变换和统计分析。通常使用的函数包括取变量的对数;计算变量的均值;判定变量的极大值和极小值;分配随机数;重新编码日期;还有生成常用分布【Distribution】的累积分布函数,如正态分布的累积分布函数。这些函数可以同算术运算合并使用。例如,"$z = \log(y + 100)$"可以通过计算 y 加 100 的和的对数,生成称作 z 的新变量。

比较算子也经常被用于转换变量或个案的选择。这些算子可以比较两个或多个变量、常量或函数。他们常用在 if-then 或 if-then-else 语句中。通常使用的比较算子包括 gt,>(大于);ge,≥(大于等于);lt,<(小于);le,≤(小于等于);eq,=(等于);ne,≠(不等于);and,&(同真);or,|(或真)。

最后,变换也常常用于从偏态【Skewed】的或非正态的分布中得到近似的单变量正态性。这有几个策略用以判断适当的变换。例如,平方根变换常常用在服从泊松分布【Poisson Distribution】的数据上。当研究者不知道数据服从的理论分布时,应当考虑幂变换(Hoaglin, Mosteller, & Tukey, 1983; Tukey, 1977)。

——Linda B. Bourque

(赵锋译校)

参考文献

Afifi, A. A., & Clark, V. A. (1996). *Computer-aided multivariate analysis* (3rd ed.). Boca Raton, FL: Chapman & Hall.

Bourque, L. B., & Clark, V. A. (1992). *Processing data: The survey example* (Sage University Paper series on Quantitative Applications in the Social Sciences, 07-085). Newbury Park, CA: Sage.

Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (Eds.). (1983). *Understanding robust and exploratory data analysis.* New York: Wiley.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

转移率(Transition Rate)

转移率,也称为"转移强度"和"转移的瞬时速率",是当研究者进行生存分析【Survival Analysis】或事件史分析【Event History Analysis】而试图区分竞争风险【Competing Risks】时,会首先遇到的一个概念。人们对于区分竞争风险的兴趣源自对死亡原因和失败原因的研究。毫不奇怪,转移率的等价旧名是"死亡率(失败)的定因力(cause-specific force for mortality)"和"定因死亡(失败)率"。因为转移率的数学定

义紧密联系于风险率【Hazard Rate】的定义，所以有些作者在两个术语间不作区分。

形式上，在某个特定的时间 t，某个事件从给定的原始状态 j 到某个目的状态 k 的转移率被定义为

$$r_{jk}(t) = h_j(t) m_{jk}(t)$$

其中，$h_j(t)$ 是时间 t 上事件从状态 j 转移的风险率，而 $m_{jk}(t)$ 是在时间 t 由某个变化构成的转移（即事件）或从状态 j 变动到状态 k 的条件概率。相似于风险率，转移率由时间的倒数（如每月或每天）测定。同样类似于风险率，转移率不能为负，但是可能超过 1。

举几个例子，如果 j 代表已婚，k 象征离异，而时间 t 指一个人的年龄，那么 $r_{jk}(t)$ 表示一个当前已婚的人在年龄 t 离婚的每年瞬时发生率。另一个例子，如果 j 表示失业，k 代表作为一名技术工人而有份工作，t 指以月计的个人失业持续期，那么 $r_{jk}(t)$ 象征一个已经失业了 t 时间长度的人作为技术工人而被雇用的每月瞬时发生率。

转移率的估计【Estimation】需要对估计任何事件发生的风险率公式作出少量调整。传统的人口学转移率估计量联系着某个具体的时间区间 $[u,v)$，被称为生命表或保险估计量【Estimator】；其定义为

$$\widehat{r_{jk}(t)} = \frac{d_{jk}(u,v)}{(v-u)\{n_j(u) - \frac{1}{2}[d_{jk}(u,v) + c_j(u,v)]\}},\quad u \leqslant t < v$$

其中，$d_{jk}(u,v)$ 是在时间区间 $[u,v)$ 上从 j 到 k 转移的数量；$n_j(u)$ 是 j 直到时间 u 还持续维持其状态的数量；$c_j(u,v)$ 是在时间 u 维持状态 j 但在时间 v 右侧被删截的数量。方程中，分母右手边乘数 1/2 来自假定转移的观测值和被删截的观测值均匀地散布在区间上 $[u,v)$。通常，现代估计量基于整合的风险率 Nelson-Aalen 估计量：

$$\widehat{r_{jk}(t)} = \frac{d_{jk}(u,v)}{(v-u)n_j(u)},\quad u \leqslant t < v$$

欲了解这些以及其他转移率估计量的数学介绍，可参见考克斯和奥克斯的著作（1984，Chap.9）。还可参见图曼（Tuma）和汉南（Hannan）（1984，Chap.3）的著作，其著作针对社会科学工作者，内容更加详细，而较少有技术的讨论。

通常，转移率用一个比例风险率模型的变体来模拟（参见 Tuma & Hannan，1984，Chaps. 5-8），

$$r_{jk}(t) = q_{jk}(t)\exp[\beta'_{jk}x_{jk}(t)]$$

其中，$q_{jk}(t)$ 或者是（在一个完全的参数模型中）一个设定的时间函数，或者（在一个部分的参数 Cox 模型中）一个未设定的冗余函数；$x_{jk}(t)$ 一个影响转移率并随时间变化的可能的解释变量集，而 β_{jk} 是表示解释变量相应影响的参数集。此类模型通常由最大似然【Maximum Likelihood】估计法或（在 Cox 模型的情形下）由部分似然估计法估值。很多时候，非参数模型也是恰当的选择，但是在经验应用中很少被使用；相关的例子，可参见吴（音译）和马丁松的文章（1993）。

——Nancy Brandon Tuma

（赵锋译校）

参考文献

Cox, D. R., & Oakes, D. (1984). *Analysis of survival data*. New York: Chapman and Hall.

Tuma, N. B., & Hannan, M. T. (1984). *Social dynamics: Models and methods*. Orlando, FL: Academic Press.

Wu, L. L., & Martinson, B. C. (1993). Family structure and the risk of a premarital birth. *American Sociological Review*, *58*(2), 210-232.

处理(Treatment)

这个主题来自实验【Experiment】设计和分析领域。这一术语有两个用法:一个是作为自变量【Independent Variable】,用以研究某个变量对一个因变量【Dependent Variable】的效应;或者,作为某个自变量的一个特定值。无论取其双重意义中的哪一个,该术语不仅过去植根于实验领域的文献,且仍将用于其中。通常,根据这个词的使用背景可以清楚地判断它是用作一个自变量还是某个变量的一个值。

另一个是作为一个变量,处理常是一个定类尺度的变量。当仅有一个实验组和一个控制组时,实验组的每个受试单元都会得到处理,而控制组的每个受试单元都不接受处理。当然,自变量的值超过两个也是可能的。如有两种不同的教学法,第一组儿童用 A 方法处理,第二组儿童用 B 方法处理,同时控制组的儿童不作处理。这时,处理即被用作定类变量【Nominal Variable】的一个值。结果变量可以是某种阅读得分,即一个定量变量。

得自此类实验的数据常用方差分析【Analysis of Variance】来分析。这也是罗纳德·费希尔先生就农业实验的设计和分析所做的开创性的工作和所处理的数据类型。

在一项实验中,处理也可以是一个定量变量【Quantitative Variable】。例如,在一项医学实验中,处理可能是药剂的注射量。第一组可能是给病人注射 5 毫克的药剂,第二组病人的剂量为 10 毫克,以此类推。同样,控制组不注射任何药剂,但有时可能注射一定量的糖水作为安慰剂。结果变量可能是某种类型的一个定量的血液检测。这时,术语处理就用上述的第二种含义,即某个自变量的特定值。一种特别的教学法是一个处理,同样,某种药剂的一个特定剂量也是一个处理。

得自后一类型的实验数据常用回归【Regression】作分析。散点图【Scatterplot】能显示数据是否拟合一个线性模型,或者其他的模型更合适。

术语处理有一个含义,即在研究中存在着某种可以被操控的事物。由于此种操控不能用于观测数据,因而在由样本调查和其他方式得到的观测数据中,无论是针对其自变量,还是自变量的值,我们都看不到这一术语的应用。

除了心理学的某些领域,实验并不常用于社会科学研究,因此处理一词也很少用在这些科学中。

——Gudmund R. Iversen

(王玥译　赵锋校)

参考文献

Box, G. E. P., Hunter, W. G., & Hunter, J. S. (1978). *Statistics for experimenters*. New York: Wiley.

Cobb, G. W. (1998). *Introduction to design and analysis of experiments*. New York: Springer-Verlag.

Moore, D. S., & McCabe, G. P. (1999). *Introduction to the practice of statistics*. New York: W. H. Freeman.

树图(Tree Diagram)

在图论中,树被定义作无环的完全连接图。像任何其他的图,一个树图由一组弧连

在一起的节点集构成。在一张树图中,有些节点是内节点,有两条或多条弧从其发散出去。其他的节点是外节点或树的“叶”,仅带有一条发散出的弧。例如,图 1(a)中,节点 A,B,C 和 D 是叶,而节点 e,f 和 g 是内节点。有时,为了构图方便,树图会以平行的形式表示,如图 1(b)所示。这时,树的弧只由图中的平行线代表,而垂直线仅用于连接弧。

在社会科学中,树图常用于呈现概念或实体间的关系。例如,树图可以用于表现某种社会网络【Network】的结构(Wasserman,1994)。可能更常见的是,树被用来表现层级聚类分析【Cluster Analysis】的结果。在这些图中,树的叶通常表示被聚类的个别实体,而树图的内节点表示实体(即聚类)之集。

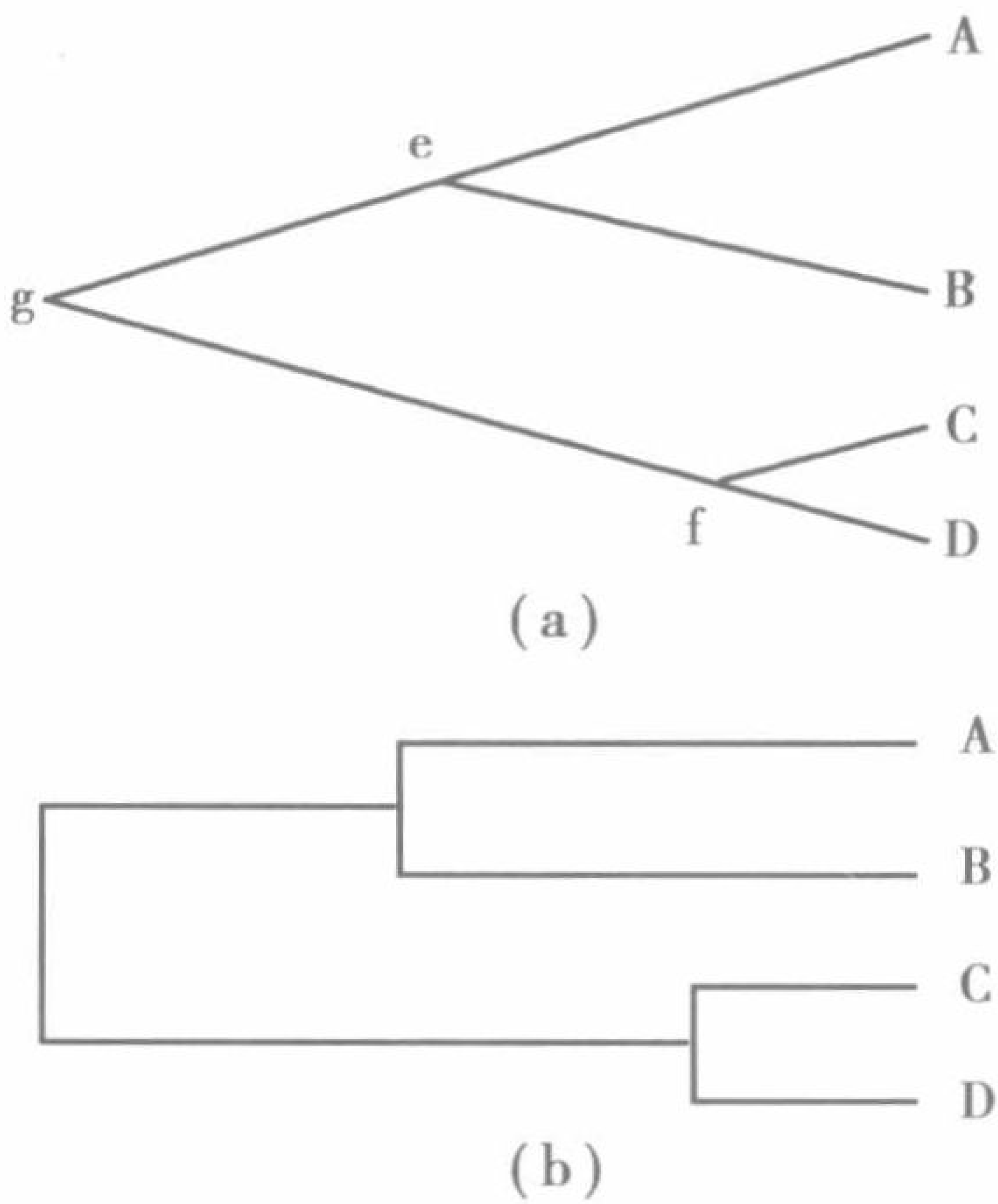


图 1 一个超公制树的两种展示

树图还特别适用于呈现实体间嵌套的层级关系。在这个世界上,此类层级关系可能以多种方式出现。当概念或实体在普遍性的几个不同水平上被归组时就可能出现。一个具体的例子,人们可以根据概念将亲属称谓归入不同的称谓类别,如核心家庭、其他直系亲属、旁系亲属(Rosenberg & Kim,1975)。当类似进化的连续分化过程出现在实体间的关系中,而这些关系又可以由一个枝出的树结构表示时,就会出现另一种层级嵌套关系,如语言的历史发展(例证,Corter,1996,见图 6.1)。

树图有时用于模型 N 个实体集合内的邻近关系,这时,树中的距离(模型距离)近似于实体对间的邻近关系(观测到的相似或不相似)。两种不同类型的树结构可用于模型邻近关系:超公制树和累加树(Corter,1996)。

图 1 中的树是一个超公制树。在一个超公制树中,距离满足超公制不等式,这表明对任何三项客体 x,y 和 v,如果 x 和 y 在树中相邻(相对于 v),则有 $d_{x,y} \leqslant d_{x,v} = d_{y,v}$。从图上看,一个超公制树的叶都等距于图上的一个定点,该点称作树的根。在图 1(a)中,节点 g 即树的根。

累加树[图 2(a)]是更一般的结构。一个累加树中的距离满足累加不等式,这表示对于任意四个客体 x,y,u 和 v,如果(x,y)和(u,v)在树结构上相对邻近,则有

$$d_{x,y} + d_{u,v} \leqslant d_{x,v} + d_{y,u} = d_{x,u} + d_{y,v}$$

这就容许累加树表现特征不同的实体。例如,在图 2 中,B 和 C 相对较长的叶茎表示,两者都相对地缺乏典型性。累加树有时如图 2(b)所示,以无根的形式呈现。

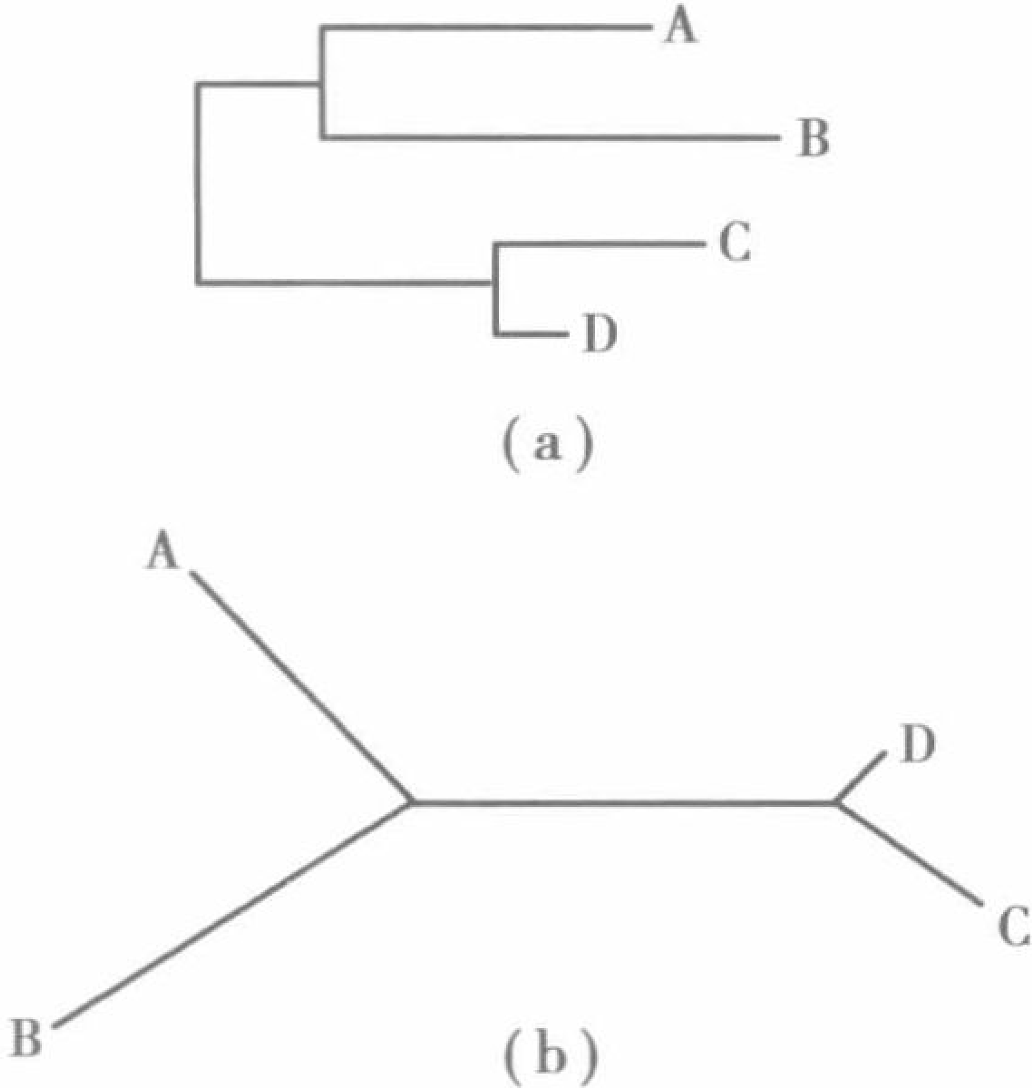


图 2 一个累加树的两种展示

——James E. Corter

(赵锋译校)

参考文献

Corter, J. E.(1996). *Tree models of similarity and association* (Sage University Paper series on Quantitative Applications in the Social Sciences,07-112). Thousand Oaks, CA: Sage.

Rosenberg, S., & Kim, M. P.(1975). The method of sorting as a data-gathering method in multivariate research. *Multivariate Behavioral Research*, *10*, 489-502.

Wasserman, S. (1994). *Social network analysis: Methods and applications.* Cambridge, UK: Cambridge University Press.

趋势分析(Trend Analysis)

许多时间序列【Time Series】变量的首要特征是它们随时间而趋向上行或下行。这一趋向可以定义为一个趋势。两个模型常用于刻画趋势数据。第一个模型是决定论趋势，$Y_t=\lambda_0+\lambda_1 T+\varepsilon_t$，其中 Y_t 是时间序列，T 是时间，λ_0 和 λ_1 是参数【Parameter】，而 ε_t 是随机的【Stochastic】震荡 $\sim N(0,\sigma^2)$。若该模型生成 Y_t，Y_t 对 T 回归【Regression】的残差则构成一个趋势平稳的过程。

第二个模型是随机趋势，$Y_t=\beta_0+\beta_1 Y_{t-1}+\varepsilon_t$，其中 Y_{t-1} 是滞后一期的时间序列，β_0 和 β_1 是参数，而 ε_t 是随机震荡。如果 $\beta_0=0$ 且 $\beta_1=1$，则该模型是一个随机步【Random Walk】；否则，就是一个带漂移的随机步。递归替代显示一个随机步等于其初值 Y_0，加上所有随机震荡之和 $\sum \varepsilon_i$。带漂移的随机步还包括 $\beta_0 T$，从而加入一个决定论部分于该过程中。随机步或带漂移的随机步的一阶差分（即 $\Delta Y_t=Y_t-Y_{t-1}$）产生一个差分固定过程。

趋势变量是非平稳的，因为它们缺少恒定的均值、方差和协方差。一个非平稳变量对其他变量的回归会危及统计推论["伪回归(spurious regressions)"]。因此，关键是发现时间序列数据中的趋势（非平稳性）。一个趋势序列当其与自身的滞后项相关时，便具有一个"特征的识别标志"。几个此类相关就构成一个自相关函数【Autocorrelation Function，ACF】。一个趋势序列的自相关在小滞后期有大的相关系数，而在更长的滞后期相关系数则缓慢地衰减。正式的统计检验（单位根【Unit-Root】检验）也广泛地用于探测非平稳性。

如果非平稳性被质疑，传统的实践是求数据的差分（假定是随机的而非决定论的趋势）。欲确定变量是否长期同时移动，则要执行协整【Cointegration】检验。协整变量可以用误差【Error】校正的形式予以模型，例如，$\Delta Y_t=\beta_0+\beta_1\Delta X_t-\alpha(Y_{t-1}-\lambda X_{t-1})+\varepsilon_t$，其中，$\Delta Y_t$ 是变量 Y 的一阶差分；ΔX_t 是变量 X 的一阶差分；$(Y_{t-1}-\lambda X_{t-1})$ 是变量 Y 和变量 X 间长期关系的误差校正机制；ε_t 是随机震荡；而 β_0，β_1，α 和 λ 是参数。α 测度了对于系统的震荡由于变量 Y 和变量 X 间协同关系而被侵蚀的速率。α 应当为负且绝对值小于 1.0。

为保证平稳性而求差分的方法的理论和统计基础都受到了评判。相关的宣称是许多社会科学的时间序列或者是"接近求和的"（随机步模型中的 β_1 略小于 1.0）或是"分数求和的"。一个分数求和的序列具有"长记忆"且可能是非平稳的。因为单位根检验的统计效力低，所以就可能无法发现接近或分数求和过程。在竞争的 ARFIMA（自回归的、分类求和的、滑动平均的）模型中，

估计分数差分参数(d)是考察接近或分数求和过程的一条有效途径。最近,协整和误差校正的概化则容许分析者为分数协整过程建模。

——Harold D. Clarke

(赵锋译校)

参考文献

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, *73*, 5-59.

DeBoef, S., & Granato, J. (2000). Testing for cointegrating relationships with near-integrated data. *Political Analysis*, *8*, 99-117.

Franses, P. H. (1998). *Time series models for business and economic forecasting*. Cambridge, UK: Cambridge University Press.

Hendry, D.F. (1995). *Dynamic econometrics*. Oxford, UK: Oxford University Press.

三角测量(Triangulation)

三角测量,指为了增强对于相继出现的研究发现的信心,使用一种以上的方法对研究问题开展调查。因为许多社会研究都是以单一研究方法的使用为基础,也因此而可能受到同这一方法相关的局限或受到这一方法具体应用的不良影响,三角测量提供了增强信心的前景。三角测量是多方法研究(multimethod research)的几项基本原理之一。这个词来自土地测量,指使用一组三角形来绘出一个地区的地图。

三角系和测量

三角测量的想法同社会与行为研究中的测量实践有紧密关联。三角测量的早期出处涉及韦布、坎贝尔、施瓦茨、西克里斯特(Webb, Campbell, Schwartz, & Sechrest, 1966)提出的非介入方法【Unobtrusive Method】的想法,他们提出,"一旦一个命题由两个或多个独立测量过程而被确认,该命题解释的不确定性就大为减少。最有说服力的证据来自一个测量过程的三角部分"(p. 3)。因此,如果我们为某个概念设计出一个新的以抽样为基础的测量,如情绪劳动,同时,根据新的测量,如果能确认由使用其他方法,如结构化观察,得到的情绪劳动的分布形态和相关关系,那么我们对新测量的信心将增大。当然,两组发现不一致的情况也是可能的,但是,正如韦布等人所说,这种情况的出现恰恰突出了仅依赖一次测量或一种方法所存在的问题。同样,两组结果不能收敛的情况还会促成新的质询路向,或者联系于相关的方法,或者联系于所涉及的实质领域。与此有关的一点是,即使三角测量的实施产生了收敛的研究发现,我们仍应当警惕作出以下结论,即这意味着研究发现是无可争论的。这可能因为两组数据都有缺陷。

三角测量的类型

邓津(Denzin,1970)拓展了三角测量的理念,使其超出了它同研究方法与设计的传统联系。他区分出四种三角测量的形式:

1.数据三角测量,即要求使用多种抽样策略来收集数据,从而获得多个人群在不同时间和不同社会条件下的数据。

2.调查者三角测量,即指在调查中使用多名研究者收集和解释数据。

3.理论三角测量,即指用多种理论立场解释数据。

4.方法三角测量,即指使用多种方法收集数据。

第四种类型,如前面的讨论所指的,是三角测量这一词最通用的意义。邓津区别了方法内的三角测量与方法间的三角测量。前者涉及使用同一方法的多种样式来考察某一研究议题;例如,一份自填问卷可能包括两个对照的量表来测量情绪劳动。方法间的三角测量涉及使用互相对照的研究方法,如问卷和观察。有时,这种意义的三角测量还包括混合使用定量研究【Quantitative Research】和定性研究【Qualitative Research】,以确定它们在多大程度上能取得收敛的研究发现。例如,休斯等人(Hughes et al.,1997)在英国的一个关于青年人对“设计师饮料”的消费研究中,同时使用了结构化访谈和焦点小组访谈。两组数据相互印证,显示出一个清晰的模式,即同这类酒精饮料的态度相关的年龄差异。

三角测量有时用来指所有使用两种或多种研究方法的情况。因此,它可能被用于指称多方法研究,即结合定量和定性研究方法,以给出一个比单独应用其中一种方法所能获得的,更加完整的研究发现。然而,有人可能论说,有理由将这个词留给那些专门的场合,即研究者寻求通过用另一种方法做交叉核验的方式,以核证研究发现的效度【Validity】。

三角测量的检验

三角测量的想法的几项基础已受到批评。首先,有人指责它赞成粗率的实在论【Realism】,即认为存在着对社会世界唯一的、确定的解释。此种实在论的立场遭到认同建构主义【Constructionism】的作者的攻击,同时,他们还论说,研究发现应当仅被视为对社会生活多种可能的译文中的一种。另一方面,在建构主义框架下研究的作者并不否认三角测量的潜力;相反,他们认为它的用处在于能够增添一项质询的丰富感和复杂感。如此一来,三角测量成为一个用以提升研究解释的可信度和说服力的工具。第二项批评是,三角测量假定,几个得自不同研究方法的数据不仅可以确定无疑地相互比较,还具有处理某一研究问题的同等能力。这一观点没能考虑到不同研究方法各自所要求的特定的社会环境,特别是那些与方法间的三角测量相关联的社会环境的差别(使用邓津的区分,1970)。例如,如果从某一结构化访谈【Structured Interview】的实施中得到的研究发现同焦点小组【Focus Group】的发现明显不合,那么,这可能是因为前者记录了私人的观点,后者记录了在焦点小组中公开的观点,所以前一观点同更一般的观点是有冲突的。

三角测量已发展出了多种意义,可是,其最常用的意义是,在一个具有收敛效度的策略下,结合使用两种或多种研究方法。近年来,由于三角测量明显赞同于一种粗率的实在论立场,它已受到诸多批评。

——Alan Bryman

(王玥译　赵锋校)

参考文献

Denzin, N. K. (1970). *The research act in sociology.* Chicago: Aldine.

Hughes, K., MacKintosh, A. M., Hastings, G., Wheeler, C., Watson, J., & Inglis, J. (1997). Young people, alcohol, and designer drinks: A quantitative and qualitative study. *British Medical Journal*, *314*, 414-418.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive measures in the social sciences.* Chicago: Rand McNally.

齐整(Trimming)

样本均值的基本问题是轻微的背离于正态性就能导致相当低的效力(Staudte & Sheather, 1990; Wilcox, 2003)。图基(Tukey,1960)首先在其研讨论文中报道了这一发现。概言之,样本方差可能完全因一个异常大或小的值(称作特异值【Outlier】)而大大地膨胀,这一结果反过来会导致在用均值而非其他可用的集中趋势量做检验时低效力的产生。

有两个一般的策略可用以处理该问题。两个策略各有其优缺点。第一个策略是检查特异值,丢弃发现的特异值,并用余下的数据计算位置量。第二个策略是丢弃固定比例的最大和最小观测值,再平均其余。这后一策略被称作齐整的均值。它不同于第一个策略,无须做是否存在特异值的经验检查。

10%的齐整均值可通过移除观测值上最小的10%和最大的10%,再平均余下的值计算得出。欲计算20%的齐整均值,则需移除最小的20%和最大的20%,再用剩余数据计算通常的样本均值。样本均值和通常的中位值都是特例,代表两个极端:无齐整和最大的量。

一个基本问题是确定应当齐整去多少数据。相对于其他可能使用的中心趋势量,要获得一个低的标准误,罗森伯格和加什科(Rosenberger & Gasko,1983)建议20%的齐整,然而当样本量小时,他们建议以25%代之。这一建议部分地源于目标——当从一个正态分布抽样时,得到一个相对小的标准误。例如,在正态性下,中位值有相对大的标准误,这反过来又意味着在检验假设时,有相对低的效力。根据人们接受的标准训练,上述建议是反自觉的,但是可以在威尔科克斯(Wilcox,2001,2003)的著作中发现其简单的解释。而且,理论和模拟显示,随着齐整的量增加,均值的问题在检验假设过程中逐渐消失。这并不是说20%总是最优解;虽然它不是,但对于一般应用而言,它是合理的选择。

观测值被齐整之后,必须使用专门的方法来检验假设(在丢弃特异值的情况下,亦如此)。此即,应用针对均值的标准方法于剩余的数据会违背用于得出这些技术的基本原则(Wilcox,2003)。

齐整的观念,用以避免特异值和非正态性的问题,也已经用在回归分析中。关于这一方法对其他稳健回归估计量的相对优点之评论已有讨论(Wilcox,2003)。

——Rand R. Wilcox

(赵锋译校)

参考文献

Rosenberger, J. L., & Gasko, M. (1983). Comparing location estimators: Trimmed means, medians, and trimean. In D. C. Hoaglin, F. Mosteller, & J.W. Tukey (Eds.), *Understanding robust and exploratory data analysis*. New York: Wiley.

Staudte, R. G., & Sheather, S. J. (1990). *Robust estimation and testing*. New York: Wiley.

Tukey, J. W. (1960). A survey of sampling from contaminated normal distributions. In I. Olkin et al. (Eds.), *Contributions to probability and statistics*. Stanford, CA: Stanford University Press.

Wilcox, R. R. (2001). *Fundamentals of modern statistical methods*. New York: Springer.

Wilcox, R. R. (2003). *Applying contemporary statistical techniques*. San Diego, CA: Academic Press.

真实分(True Score)

在测量理论中,人们设想存在着真实分,并假定每个被测变量 X 由两部分构成:应答者的真值或真实分,以及随机误差。此即,$X=T+e$。这一概念与信度【Reliability】联系在一起。一个没有随机误差的度量,同它的真实分完全一致,即有完美的信度。相反,一个仅有随机误差的度量,同它的真实分完全不一致,则具有零信度。

——Tim Futing Liao

(赵锋译校)

删节(Truncation)

参见截删和删节【Censoring and Truncation】。

真确准则(Trustworthiness Criteria)

真确准则给出了一些衡量标准,而通过这些标准,所用定性研究方法和发现的严格性、有效性【Validity】和系统性质可以得到评判。该名称所包含的准则适用于自然的、解释的、非实证的和非实验的质询模式,而这些质询模式常常被标志作建构主义的、批判理论、行动质询、现象学个案研究【Case Study】,或者自然主义范式。这些准则最初被提出(Guba, 1981; Lincoln & Guba, 1985)用以回应一系列严肃的质疑,即根据新的范式模式定性研究的合理性、可信性【Reliability】和严格性。

严格性的传统(实证主义的)准则已得到很好的确立,且易于理解。它包括:内在效度【Internal Validity】、外在效度【External Validity】、信度(可重复性)和客观性【Objectivity】。这四个严格性范畴分别同研究发现的现实性、概括性和稳定性(也即任何给定研究可被重复的能力)以及避免调查者偏倚具有同等结构。然而,现象学的或是新范式的质询明确地拒绝任何这四个范畴中的任何一个,认为它们既难以达到(无论是实践上,还是根据哲学理由),也是不适宜的(基于范式的理由)。由此,人们提出了四项替代的准则,更加对应于现象学质询的公理要求和方法论要求,同时,研究设计的策略则更适宜于那些常常,但不是总是发生于自然场景下的定性研究、定性评估或政策分析。

为了创造用于评判非定量的和自然-范式的质询的标准,早期提出了真确准则:确实性(credibility)、移植性(transferability)、可靠性(dependability)和验证性(confirmability)。这些准则被当作"真确准则",对应于传统研究的四个准则。对"真确准则"的关注显示出人们对研究发现的严格性颇具信心。这些被设计出并平行适用于实验模式的准则还被批评,被认为是"基础的",因为它们同源自传统质询的基础相联系,而不是源于现象学(建构主义的)质询的内在特征。如同传统的严格性准则,这些准则首先关涉方法论的考量;因此,在考察个案研究的方法论的适当性时,这些准则仍旧是非常有用的。

确实性指一项说明(一项个案研究)的似真性:它可信吗？报告是否被那些具有报告的生活经验【Lived Experience】的人认为是一项准确的陈述？移植性指报告能够被用于(移植到)其他同原来个案发生的场景相似的情形的能力。然而,在某个重要的方面,对移植性的判断不同于对概括性的判断。在传统的研究中,概括性的估值由研究者做出。但是,在估计移植性时,结果有用性的判断由"容纳的背景"给出(Guba & Lincoln, 1989; Lincoln & Guba, 1985),比如研究发现的使用者和消费者,以及那些希望在自己的情境中使用发现的个体。可靠性指所用研究过程的稳定性、可追溯性及其内在的逻辑。通过对研究过程的审查,即分析方法论决策的制定,并证明其合理性来评定可靠性。验证性证明个案中报告的数据完全可追溯到原始数据源(比如,田野笔记【Fieldnotes】、档案、原始记录、观察日志等)。

真确准则首先是方法论的准则,它们回应的是方法[尽管某些建议的方法(Guba & Lincoln, 1989)——直接联系到现象学前提,比如通过成员审定来查明研究者是否已经理解回答者的社会建构],而非直接同范式假定相关联。然而,因为新的范式质询常需要非传统的、定性的方法,还因为真确准则同这些方法的使用紧密地联系在一起,所以它们在判定突显-范式质询(emergent-paradigm inquiry)的基本元素和方法的严格性(如内容的吻合和持续的水平)上仍然有效。

——Yvonna S. Lincoln

(赵锋译校)

参考文献

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communications and Technology Journal*, *29*, 75-92.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

t 比率(*t*-Ratio)

参见 *t* 检验【*t*-Test】。

t 统计量(*t*-Statistic)

参见 *t* 分布【*t*-Distribution】。

t 检验(*t*-Test)

威廉·戈塞特在位于都柏林的 Guinness 啤酒酿造厂从事质量控制时,以笔名"学生(student)"写作了多篇数理统计学文章。在他 1908 年题为"均值的可能

误差”中给出了学生 t 检验之基础。将他的学生 t 分布【*t*-Distribution】用在许多实际应用中对于分析数理统计学中的小样本是一项重要的贡献。

单均值的学生 t 检验

此类应用的例子之一是单个均值的检验。当 σ 未知且样本量大时，我们可以有把握地说，均值的估计标准误 $\hat{\sigma}_{\bar{X}}$（满足 H_0 时，均值的抽样分布之标准差）将等于其真值 $\sigma_{\bar{X}}$，同时，我们评估的比值，参照正态抽样分布，有标准分 $z=[\bar{X}-E(\bar{X})]/\sigma_{\bar{X}}$。然而，当样本小时，我们不能对我们的标准误估计抱有信心；此外，我们所要用的比值不再是 z 分数，而是 $t=[\bar{X}-E(\bar{X})]/\hat{\sigma}_{\bar{X}}$，公式中的分母不再是常数，而是一个随机变量，它随样本量而变化，亦即随自由度【Degrees of Freedom】变化。

t 分布呈均值为 0 对称且单峰的形状，但是与正态分布【Normal Distribution】相比，其中部较平，两端更密。随样本量 n 增大，t 分布趋向标准正态分布。

现在，我们假设某个特殊群体，如化学家的平均收入为每月大于 2 000 欧元（H_a：$\mu>2\ 000$），不同于虚无假设 $H_0:\mu=2\ 000$，且标准差未知。我们抽取了一个 $n=26$ 的随机样本【Random Sample】，得到 $\bar{X}=2\ 040$，且 $s=100$。我们对 $\hat{\sigma}_{\bar{X}}$ 的估值为 $\frac{\hat{\sigma}}{\sqrt{n}}=\frac{s}{\sqrt{n-1}}$，由此，$t$ 比值为 $t=\frac{2\ 040-2\ 000}{100/\sqrt{25}}=2$。根据 t 表，我们查询单侧检验且 $\alpha=0.05$，得到 t 的临界值 $t^*=1.706$。我们的样本值大于临界值，此即，化学家的平均收入显著地高于 2 000。即便样本量如此小，我们仍可得到显著的结果。然而，如果我们实施双尾检验（$t^*=2.056$），那么结果将不再显著。

均值差的 t 检验

对学生 t 检验的另一种应用是均值差【Difference of Mean】的 t 检验。我们可以设想一实验设计中有两个群体，例如，一组男性和一组女性，就这两组考量其在某个定量变量【Quantitative Variable】上的得分，比如说他们的吸烟行为，可由每天吸烟的数量测量。在此例中，因变量 Y（吸烟量）是连续的【Continuous】，而自变量 X（代表两个不同的组）是二分的【Dichotomous】。我们可以给试验组和控制组各自编码作 $X=1$ 和 $X=2$。我们想考察是否男性组（$X=1$）的平均吸烟数显著地不同于（双尾检验）或显著地少于（单尾检验）女性组（$X=2$）的平均吸烟数。虚无假设陈述均值差等于零。两组各取 5 人，即 $n_1=n_2=5$。设总样本的得分序列为 2 2 3 4 4 5 6 7 8 9，均值为 5，方差为 54。男性组的得分为 2 2 3 4 4，均值为 3，方差为 4。女性组的分数是 5 6 7 8 9，均值 7，方差 10。两组散布程度不同：方差分别为 4 和 10。欲合理地实施 t 检验，分布应当近似正态，且方差必须没有显著不同。就这两个条件而言，有两项检验程序，正态性检验和同方差检验【Levene’s Test】，分别对应于上述要求。假定两个条件都满足，可计算如下：

两方差的均值（合并的平均数）之计算：

$$s_w^2=\frac{\text{组 1 的变异}+\text{组 2 的变异}}{(n_1-1)+(n_2-1)}$$

$$=\frac{4+10}{(5-1)+(5-1)}=1.75$$

欲获得抽样分布，我们需考虑所有理论可能的随机且独立的样本对（各自的样本量为 n_1 和 n_2），这些样本对都抽取自一个理论总体（即 H_0 之下的总体，就其而言，两样本均值差为零或至少纯然随机）。对无数个样

本对的每一个,我们考量吸烟数的均值差。据此,得到一个均值差的抽样分布。该抽样分布的方差似乎等于两个独立的抽样分布的方差之和(其中一个样本量为 n_1,另一个样本量为 n_2)。由此,

$$\sigma_{\text{diff}}^2 = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}$$

公式中的 σ^2 项,即理论总体(根据 H_0,两个样本得以从中抽取)的方差未知,因此,需要估计而得。使用的估计量是合并的平均数 s_ω^2,如上计算。为了能够应用统计表,我们要实施标准化:对于每一个均值差,需要减去均值差的抽样分布的均值(根据 H_0,这里等于0),其结果被标准误 σ_{diff}除(其值等于抽样分布的估计方差的平方根)。由此,我们获得 t 分数而非 z 分数;其理由是样本量小,进而总体方差必须被估,这又源于对正态分布的背离。如果这里就我们的样本对计算 t 分数,就可以将之置于抽样分布之中,从而得出结果:

$$t = \frac{|\bar{Y}_{(1)} - \bar{Y}_{(2)}| - 0}{\left(\frac{s_\omega^2}{n_1} + \frac{s_\omega^2}{n_2}\right)^{1/2}} = \frac{|3 - 7| - 0}{\left(\frac{1.75}{5} + \frac{1.75}{5}\right)^{1/2}} = 4.78$$

针对先前假设的一类错误,$\alpha = 0.05$,以及自由度 $(n_1 - 1) + (n_2 - 1) = (5 - 1) + (5-1) = 8$,并应用单尾检验(即猜想当下女性吸烟多于男性),我们在 t 表上找到临界值 $t^* = 1.86$。我们已计算的 t 的绝对值更高,因而,在虚无假设成立的情况下有低于5%的出现概率,亦即,结果位于拒绝域之内。于此,以大于 95%不会犯错误的可能性,女性和男性平均的吸烟行为间存在着显著的差异,即女性吸烟多于两性的平均数。

一些考量

这里使用的均值差检验的例子与条目中关于 F 比值的例子一样,此外,读者可以证明那里得到的 F 值 22.86(计入舍入误差)等于查到的 t 值 4.78 的平方,因为当自由度为 1 时,$F=t^2$。

针对两个均值间差的检验的虚无假设表明其差为零。当然,也可使用其他一些非零的数值,还可以假设均值差不同于那个数值。

当多于两个样本被考察时,就不容许用几个 t 检验的办法来比较所有的组别;因此,必须实施一个 F 检验。如果我们真想实现被称为对比检验的多元比较,那么就必须更保守地做检验,此即,以更低的第一类错误【Type Ⅰ Error】的概率(参见方差分析【Analysis of Variance】和相关的话题)。

读者不应忘记 t 分布只能被用在当总体分布是正态的要求被满足的情况下。这点不仅对单个均值的检验有效,对均值差的检验也有效。在大多数情形下,没有关于总体分布的信息,所以必须对样本做正态性检验。

在我们的例子中,假定方差齐性。然而,这一假定可以由一个 F 检验验定。如果这项检验论断方差 s_1^2 和 s_2^2 间有显著差异,那么,计算两个方差的合并平均数 s_ω^2(用来作为 σ^2 的估计值)的方法必须代之以从每个样本中计算各自标准误的方法。然后,自由度的适当数值也必须予以校正。此方法的公式可以在某些教科书中找到,如海斯(Hays,1972)和布莱洛克(Blalock,1972)的著作。

应当说明的是,例子中的两个样本必须是独立样本。当观测值是配对的,如当个体由研究者配对时,或者,假设组 1 中的男性和组 2 中的女性是夫妻时,必须选用不同的方法。这时,我们的陈述 $\sigma_{\text{diff}}^2 = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}$不再

有效,但是一个协方差可能会加入 σ^2_{diff} 的计算中。相应的方法把数据看作来自配对总体的一个样本,进而实施单个均值的 t 检验。

最后的一项评论是 t 检验可用于更高级的分析,如用于多元回归【Multiple Regression】的参数显著度检验。这时,也可由一个 F 检验代替,而 F 将是 t 的平方。我们也要提一下霍特林发展了 t 的高级版本,即 Hotelling T^2 检验,它可用于针对两个群体和多个变量的判别分析【Discriminant Analysis】。

——Jacques Tacq

(赵锋译校)

参考文献

Blalock, H. M. (1972). *Social statistics*. Tokyo: McGraw-Hill Kogakusha.

Hays, W. L. (1972). *Statistics for the social sciences*. New York: Holt.

Student. (1908). *The probable error of a mean*. In E. S. Pearson & J. Wishart (Eds.), *Student's collected papers*. London: University College.

Tacq, J. J. A. (1997). *Multivariate analysis techniques in social science research: From problem to analysis*. London: Sage.

二十陈述测验(Twenty Statements Test)

自我和认同是心理学和社会学共同关心的两个概念,因为它们让研究者得以理解人们如何把自己放在不同的社会类别中,以及如何反思他们是谁。二十陈述测验(有时称作"我是谁"测验)是多种用来测量自我概念的工具之一。它由符号互动论社会学家曼弗雷德·库恩和托马斯·麦克帕兰(Kuhn & McPartland, 1954)引入,其目的在于使乔治·赫伯特·米德(一位主要的实用主义哲学家)的自我概念更加严密和操作化。

在回答"我是谁"的问题时,受试者被要求写下首先进入他脑海中的 20 个句子。他们还被告知答案没有正确或错误之分。测验是一个开放式的工具,其中既可列出名词,也可列出形容词。在库恩和麦克帕兰的早期工作中,他们预测随着年龄的增加以及社会变迁,比起实际的名词回答(例如,妻子、学生、男人)、形容词回答(例如,兴高采烈的、虔信的、意识到存在的)可能更频繁地被使用。

自我理论特别有兴趣了解自我的观念如何变化。在传统世界中,"前现代的认同"可能通常被视为当然。人们作为世界的一份子,通过传统和宗教的确认,带着同他人分享和共有的身份,知道他们自己是谁。这种身份不仅不被视为问题,还常常被视为当然。相反,随着现代性的开展,人类个体和个体的身份认同变得更加重要。逐渐地,关于一个人是谁的自我反观成为一个问题。二十陈述测验即是用以研究此类问题的一个有价值的研究工具。

这个工具是众多的测量自我概念的工具之一,不过近年来却少有使用。尽管它容许对某种自我感作广泛的、概略的描述,但它太基础了。此外,对这种方法的批评认为它没有考虑到:(a) 自我的情境性,即自我随处境的变化而改变;(b) 重要他人在影响所陈述内容中扮演的角色;(c) 编码和内容分析的问题,而这些问题通常与开放式问卷的问题有关。

——Ken Plummer

(王玥译　赵锋校)

*也可参见自陈式测量【Self-Report Measure】、符号互动论【Symbolic Interactionism】。

参考文献

Bracken, B. (Ed.). (1995). *Handbook of self concept: Developmental, social and clinical consequences.* New York: Wiley.

Hewitt, J. (1989). *Dilemmas of the American self.* Philadelphia: Temple University Press.

Kuhn, M., & McPartland, T. S. (1954). An empirical investigation of self attitudes. *American Sociological Review*, *19*, 68-76.

Wylie, R. (1974). *The self concept: A review of methodological considerations and measuring instruments.* Lincoln: University of Nebraska Press.

双侧检验(Two-Sided Test)

参见双尾检验【Two-Tailed Test】。

二阶段最小二乘(Two-Stage Least Squares)

二阶段最小二乘【TLS 或 2SLS】是一般用于估计联立方程【Simultaneous Equation】模型的参数【Parameters】的程序。当面临互为因果时,给定一般联立方程模型,其中变量相互影响,普通最小乘方【Ordinary Least Squares】将产生偏倚【Biased】的估值。然而,二阶段最小二乘,一个工具变量【Instrument Variables】的方法将在此情形下产生一致的参数估计【Estimation】。设想下列模型:

$$Y_1 = b_0 + b_1X_1 + b_2Y_2 + e \qquad (1)$$

$$Y_2 = a_0 + a_1Y_1 + a_2X_2 + u \qquad (2)$$

TLS 估计循阶段而推进,在一个时刻估计一个方程的参数。让我们估计方程(1)。在第一阶段,系统中所有的外生变量(X_1,X_2)【Exogenous Variable】成为自变内生变量(Y_2)【Endogenous Variable】的回归项。这就构造了一个工具变量,预测的 Y_2 它完全是外生变量的函数。在第二阶段,预测的 Y_2 代替观测的 Y_2 进入方程(1),且该方程以普通最小乘方法估计。这些第二阶段的估计值是一致的,因为方程中所有自变量【Independent Variables】是有效地外生的。相同的二阶段程序再用于方程(2)以获得方程的参数估值。

——Michael S. Lewis-Beck

(赵锋译校)

*也可参见因果模型【Causal Modeling】、工具变量【Instrumental Variable】。

双尾检验(Two-Tailed Test)

虚无假设【Null Hypothesis】经常用一个特定的值设定总体参数【Parameter】,即便我们不必然相信这个值就是参数的真值。在研究两个变量间的关系时,用回归【Regression】分析做例子,最常用的虚无假设是两个变量间不存在关系。这可以翻译

为虚无假设 $H_0:\beta=0$,其中 β 是总体回归线的斜率【Slope】。最常见的是,我们为了证明两个变量间存在关系而研究两个变量间的关系【Relationship】。而如果我们能够拒绝关于不存在关系的虚无假设就能实现这一目标。

如果我们能显示总体回归线的斜率不等于零——通过拒绝虚无假设———那么这一斜率有一正值还是负值呢?如果我们没有关于斜率任何额外的信息,那么替代的假设可写成 $H_a:\beta\neq0$,意味着斜率可以是负,也可以是正。这样一个替代假设使得检验成为一个双尾【Two-Tailed】检验,也称双侧【Two-sided】检验。

在最初处理该问题的方法中,如罗纳德·费希尔爵士和其他人所发展的,他们选择一个预设的显著水平【Significance Level】,比如说,$\alpha=0.5$,因为缺少更好的值。依据分析的性质,他们选择相应的统计量,常用的有 Z,t,卡方【Chi-Square】或 F。就正态 Z 变量而言,2.5%的值小于-1.96,同时,2.5%的值大于1.96。因此,对于正态 Z 变量,如果观测到的 Z 值小于-1.96或大于1.96,则拒绝虚无假设,意味着该值落入分布的两个尾部之一。

这意味着对于一个真的虚无假设重复研究许多次,我们将得到一组样本,并有一组 z 分数,其中的5%或者小于-1.96,或者大于1.96。就这些样本而言,我们可能错误地拒绝虚无假设。之所以名为“双尾检验(two-tailed test)”是因为通常较大或较小的检验统计量 z 或 t 才会拒绝虚无假设。根据卡方检验和 F 统计量检验的性质,两者都被设定为双尾检验。

由于使用统计软件来分析数据,于是就不再从先于分析选择的置信水平 α 开始检验,而是使用作为分析的一部分计算出的 p 值【p Values】开始。就 z 和 t 而言,如果我们见到绝对值符号 $|z|$ 或 $|t|$,那么相应的 p 值代表双尾检验。如果我们开始分析前决定做双尾检验,而软件给出一个单尾的 p 值,那么我们必须加倍该 p 值。

当然也可以只是报告算得的 p 值,而让读者确定检验是双尾的,抑或单尾的。

——Gudmund R. Iverson

(赵锋译校)

参考文献

Mohr, L. B. (1990). *Understanding significance testing* (Sage University Paper series on Quantitative Applications in the Social Sciences, 07-073). Newbury Park, CA: Sage.

Morrison, D. E., & Henkel, R. (Eds.). (1970). *The significance test controversy*. Chicago: Aldine.

双因素方差分析(Two-Way ANOVA)

双因素方差分析也即方差的双因素分析,是一种用于分析来自特定研究数据【Data】的检验,在此种研究中,调查者希望检验两个变量【Variable】对某个行为的量的独立和联合的效应。数据来自因子实验或研究,其中两个变量或因子(如因子 A 和因子 B)的每一个独立的水平都被识别出且所有的组合也被构造出。因子 A 和因子 B 可能是被操控的自变量【Independent Variable】,如动机的水平或在小组中个人表现的状况,或者是主体变量的被测水平,如年龄和性别。那些在社会科学中更令人感兴趣的应用常包括一个被操控的自变量因子 A 和一个主体变量因子 B。主要兴趣在于,因子 A 对响应变量是否随因子 B 的不同水平而有所不同。

例如,考虑这样一个研究,由于受到暴力视频游戏的感染对年轻的男孩和女孩随

后发起攻击行为的效应。让我们设定被编码的行为——因变量【Dependent Variable】——是一个孩子在15分钟的嬉戏期间打、踢或推其他孩子的次数。以某个年龄的儿童为总体(设6~7周岁),孩子们被随机地分入一种情形,其中他们在同其他孩子嬉戏前,一部分先玩30分钟的暴力视频游戏,另一部分玩30分钟无暴力的视频游戏。这构成了对自变量的操控。针对男孩和女孩分别实施随机分配【Random Assignment】,于是得到一个2×2的设计,其中两个视频游戏类别的水平同两个性别水平相交叉。

由此得到的数据和双因素方差分析可以用于回答下列研究问题:(a)男孩比女孩更有攻击性吗?(b)有了暴力视频游戏的经验,儿童变得更有攻击性吗?(c)暴力视频游戏的效应在男孩和女孩上不同吗?在方差分析的用语中,主效应【Main Effect】检验表述了前两个问题,即根据实验条件做平均,平均攻击得分对男孩而言显著地高于女孩吗?以及根据性别做平均,暴力视频游戏的平均攻击分显著地高于无暴力视频游戏的平均攻击分吗?第三个问题,可能是最有兴趣的问题之一,由交互效应【Interaction】的检验表述——在男孩和女孩之间,暴露于视频游戏的效应有显著的区别吗?

展示结果

这三个检验,即两个主效应和一个交互效应,在双因素方差分析中是相互独立的,而这意味着任何显著的和不显著的结果之组合都是可能的(相应的计算程序可参见Hays,1994;Jaccard,1998;Levin,1999)。一个ANOVA表,配之以因子A和因子B的每种组合的均值结果图,给出了完整的景象。上面例示了两个有趣的、假设的结果。在第一个例子中,方差表显示两个主效应都显著,而交互效应不显著(表1)。图1显示男孩和女孩有近似平行的线,说明虽然男孩比女孩更有攻击性,同时男孩和女孩暴露于暴力视频游戏同暴露于无暴力视频游戏相比,攻击行为的数量有相似的增加(图1)。在第二个例子中,两个主效应都再次显著且交互效应也显著(表2)。

表1 显著的主效应同不显著的交互效应的情形

类型	*DF*	*SS*	*MS*	*F*	*p*
性别	1	122.51	122.51	33.56	<0.000 1
游戏	1	103.51	103.51	28.35	<0.000 1
性别×游戏	1	1.01	1.01	0.28	<0.600

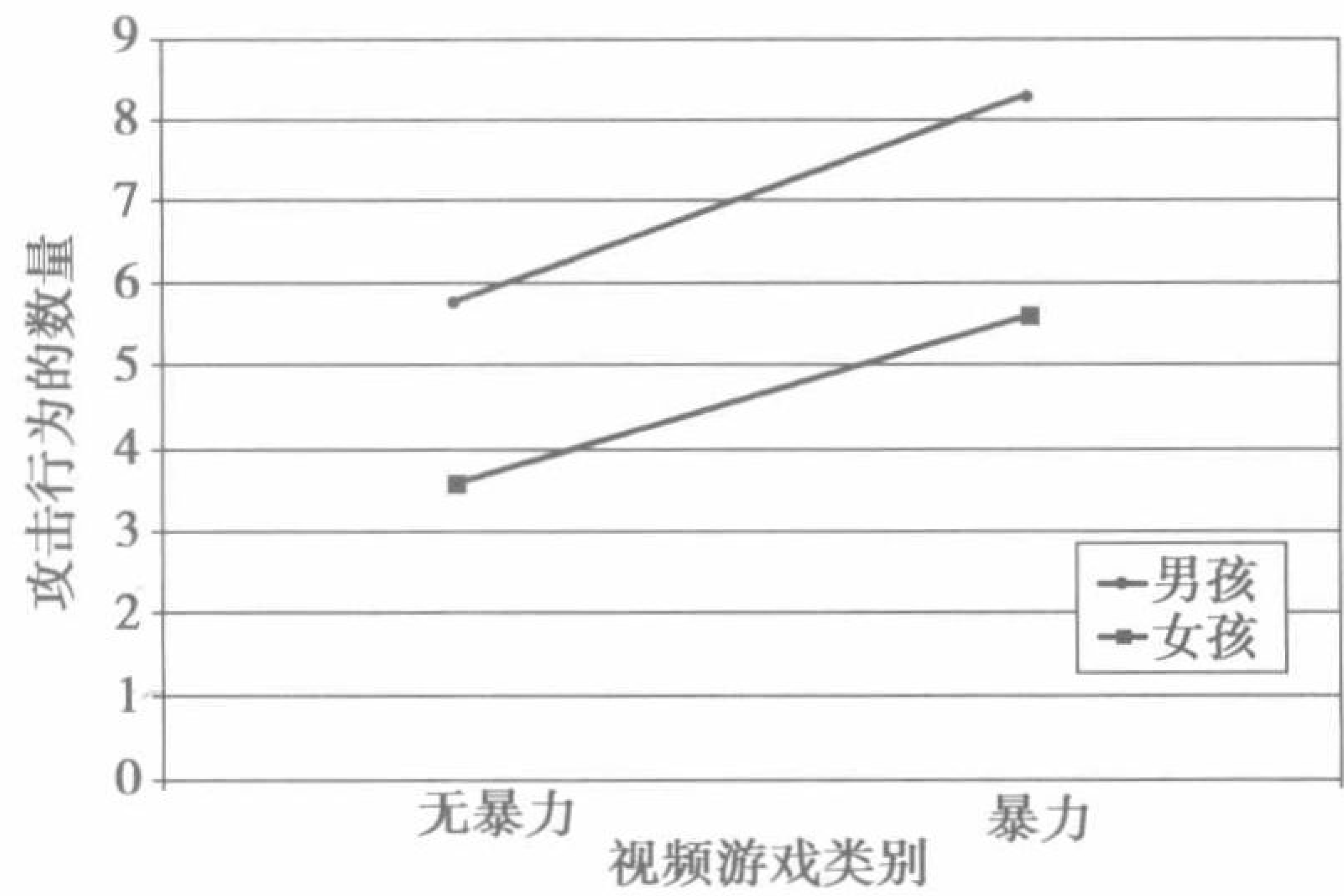


图1 攻击行为作为性别和视频游戏类别的函数

表2 显著的主效应同显著的交互效应的情形

类型	*DF*	*SS*	*MS*	*F*	*p*
性别	1	137.81	137.81	50.76	<0.000 1
游戏	1	103.51	103.51	38.12	<0.000 1
性别×游戏	1	37.81	37.81	13.93	<0.000 4

相应的图显示,两条线在暴力端的间隔大于它们在无暴力端的间隔(图2)。这表明,男孩比女孩受到暴力视频游戏更强烈的影响。探明和描述交互效应能让我们区分两种结果的区别内容。结合因子研究,应用双因素方差分析容许研究者在一个单项研究中考察两个变量分别的和联合的效应。相同的原则可以扩展至有三个变量的三因素方差分析,以及更多维的情况。

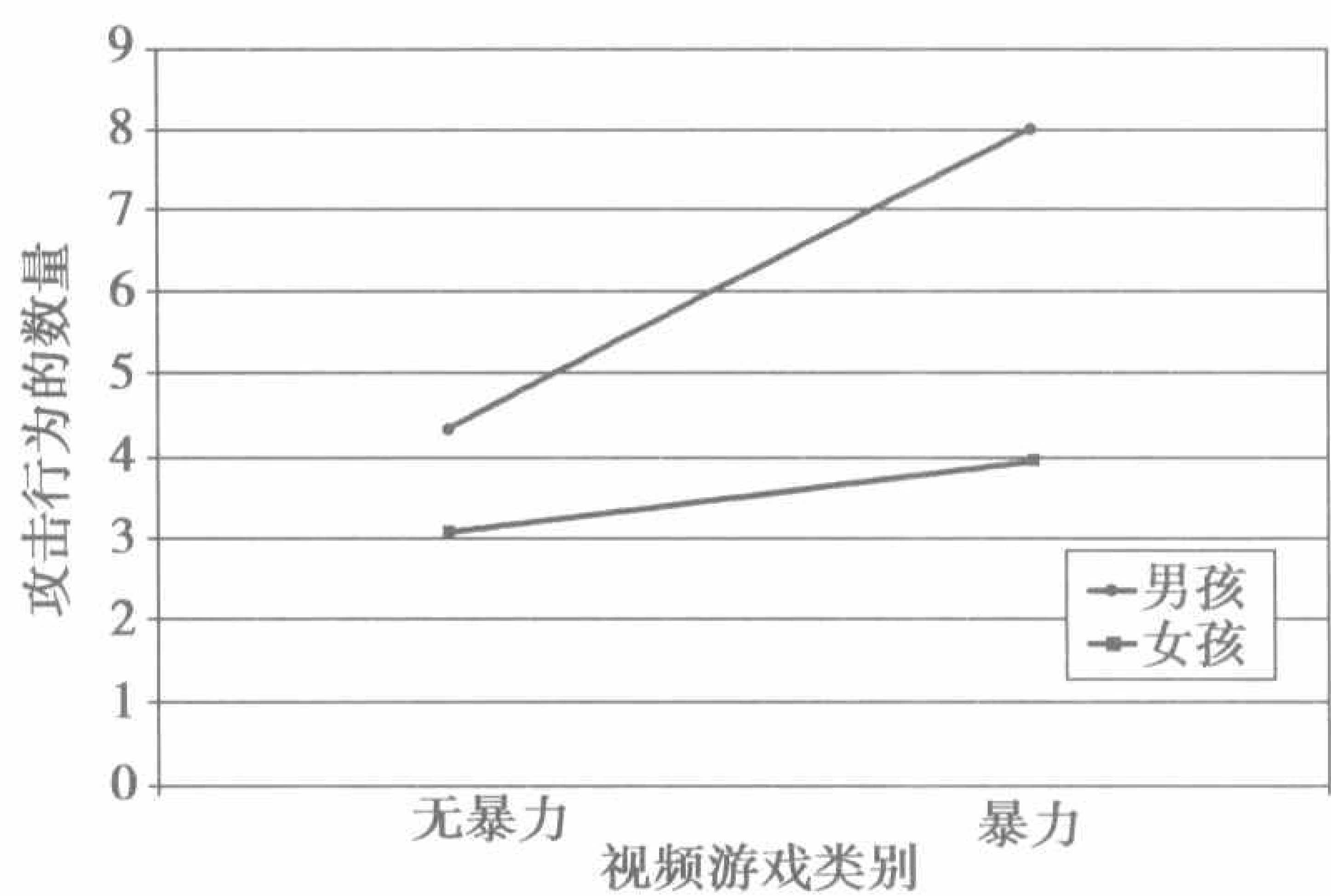


图 2 攻击行为作为性别和视频游戏类别的函数

——Irwin P. Levin

(赵锋译校)

参考文献

Hays, W. L. (1994). *Statistics* (5th ed.). Fort Worth, TX: Harcourt Brace.

Jaccard, J. (1998). *Interaction effects in factorial analysis of variance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-118). Thousand Oaks, CA: Sage.

Levin, I. P. (1999). *Relating statistics and experimental design: An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-125). Thousand Oaks, CA: Sage.

第一类错误(Type Ⅰ Error)

第一类和第二类错误是可能发生于假设检验【Hypothesis Testing】中的错误。即便数据只是对随机抽样【Random Sampling】的观测值有效，而不是对于整个总体【Population】的观测，样本数据仍然可用于作出关于更大总体的推论。这意味着从小样本到大总体的外推，因此，我们可能犯错误。有两类主要的错误，分别称为第一类错误和第二类错误。

从样本至总体的外推可由虚无假设【Null Hypothesis】完成。我们可以设定一个或多个值作为总体参数。这里有几个关于虚无假设的例子：H_0 ：μ = 100(总体均值【Mean】等于 100)；H_0 ：$\sigma^2 \geq 25$(总体方差【Variance】大于等于 25)，H_0 ：$\beta=0$(总体的回归系数【Regression Coefficient】等于 0)。我们想要了解总体参数的值，但由于这些值是未知的，虚无假设就给问题以额外的信息。就第一个虚无假设而言，总体均值可能等于 100 吗?

考虑这样的情形，即虚无假设实际是正确的。如果样本数据导致我们拒绝虚无假设，并得出它非真的结论，那么我们就犯了一个错误。因为虚无假设实际为真，所以它不应当被拒绝。这就是第一类错误，即便虚无假设实际为真，仍然拒绝它。

我们根本不知道是否犯了这一错误，因为总体参数的真值永远保持未知。我们所能做的仅是，在虚无假设为真且数据拒绝虚无假设时，去查看有何事发生。一个可能是猜想我们会多频繁地犯第一类错误。设若，终其一生，我们检查 1 000 个不同的虚无假设，其中 800 个是错误的，另 200 个是正确的。如果有 5%的时间犯第一类错误，那么在 200 个真假设案例中，就可能有 10 个案例犯了错误(10/200 是 5%)。另外，800 个虚无假设与此无关，因为我们仅处理虚无假设为真的案例。

欲确定我们是否拒绝一个虚无假设，要使用样本数据来计算所谓的 p 值【p Value】。大多数统计软件会为我们计算这一数值。它告诉我们，在虚无假设为真的情形下，有多频繁地获得我们的样本数据或者更极端的数据。例如，设 H_0 ：$\beta=0$，且根据数据 $b=4.56$，相应的 $p=0.001$。因此，如果总体回归线真正地水平(两变量间没有关系)，我们在 1 000 次重复的该研究中，仅会得到一次回归系数为 4.56 或更大系数的样本。

——Gudmund R. Iversen
(赵锋译校)

参考文献

Iversen, G., & Gergen, M. (1997). *Statistics: The conceptual approach*. New York: Springer-Verlag.
Moore, D. S., & McCabe, G. P. (1998). *Introduction to the practice of statistics*. New York: W. H. Freeman.
Utts, J. M. (1996). *Seeing through statistics*. Belmont, CA: Duxbury.

第二类错误(Type Ⅱ Error)

统计推论【Statistical Inference】方法被用于推论未知参数【Parameters】的值。在假设检验【Hypothesis Testing】中,一个虚无假设【Null Hypothesis】被设立,其中设定一个或多个关于参数的值:H_0 : $\mu=100$(总体均值【Mean】等于100),H_0 : $\sigma^2 \geq 25$(总体方差【Variance】大于等于25)或H_0 : $\beta=0$(总体回归系数【Regression Coefficient】等于0)。因为参数值未知,所以虚无假设成为关于参数的可能值的陈述。这如同我们询问,参数是否会取如虚无假设所设定的值。

样本【Sample】数据被用于回答虚无假设所提出的问题。根据数据,我们作出决定,拒绝还是不拒绝虚无假设。因为这一决定是从有限样本数据至更大总体的外推,所以作决定时可能犯错误。如果实际的参数值是虚无假设所设定的值,那么虚无假设为真。若我们仍然拒绝虚无假设,则我们犯下第一类错误【Type Ⅰ Error】。

然而,也可能实际的参数值不等于虚无假设所设定的值。在这种情形下,虚无假设是错误的,而且应当被拒绝。但是,如果数据导致我们没能捕捉到虚无假设为伪的事实,那么我们就犯了错误。这是第二类错误——没有拒绝一个错误的虚无假设。

是否犯此类错误取决于两件事:第一,取决于参数的真值距离虚无假设的设定值有多远。如果真正的总体参数均值$\mu=101$,而虚无假设陈述总体均值$\mu=100$,那么虚无假设是错误的且应当被拒绝。但是由于两者间的差值太小,这就很难发现虚无假设是错误的。在多次重复的研究中。样本均值将聚集在101周围,而非100周围。即使这样,这种差异可能小到对我们手头的问题无关紧要的地步。

第二,数据是否能发现虚无假设为伪还取决于样本中观测值的多少。若是小样本,就比大样本更难发现虚无假设为伪;若是大样本,在多次重复的研究中,样本统计量将更紧密地聚集在真正的参数值周围,进而可以更频繁地拒绝虚无假设。就一个国家的参数和给定的样本量而言,我们可以计算,得到一个虚无假设被错误拒绝的数据的频繁度(概率【Probability】)。检验的效力(power)即被定义为一减此概率。

——Gudmund R. Iversen
(赵锋译校)

参考文献

Iversen, G., & Gergen, M. (1997). *Statistics: The conceptual approach*. New York: Springer-Verlag.
Moore, D. S., & McCabe, G. P. (1998). *Introduction to the practice of statistics*. New York: W. H. Freeman.
Utts, J. M. (1996). *Seeing through statistics*. Belmont, CA: Duxbury.

U

非平衡设计(Unbalanced Designs)

非平衡设计(有时称为非正交设计)是与定性自变量【Independent Variable】相关的因子【Factor】研究设计。这一情况很像回归【Regression】分析中的多重共线性【Multicollinearity】,难以对相互关联的自变量的效应作出准确的估计。一般而言,当样本量不相等时,即每格中受试者的数量不等时,设计就是不平衡的。其特例是存在空格。

叶芝(Yates,1934) 证明,在分析来自非平衡设计的数据时,就大多数目的而言,通常的方差分析【Analysis of Variance, ANOVA】会引致错误。使用类型Ⅰ平方和【Sum of Squares,SS】,ANOVA 会根据变量在模型中的次序,将解释权重分给各个自变量。第一个变量会得到所有与之混合的方差,无论这个方差是该变量单独具有的,还是与其他某个或多个自变量共同享有的。位于第一个自变量后的自变量仅会被给予一个额外增加的解释权重,后序的变量依次如此。

为了解决 Type Ⅰ SS 的这种任意性,叶芝发现了一种被称为加权均值平方和的方法,现在一般称为 Type Ⅲ SS,而这种方法已成为处理非平衡数据的常规分析选择。在这种方法中,每一个自变量都被当作一个 Type Ⅰ模型中最后的自变量,被分予平方和;此即,一个变量的权重是当其他因子的解释权重都被考虑过后,由其对整个解释掉的方差的唯一贡献所确定。因此,一个自变量的平方和实际上是它的偏平方和。请注意,与其他自变量混合在一起的平方和在这里被遗留下来,未予分配。

内尔德和莱恩(Nelder & Lane,1995)认为,这个方法是不切实际的,因为除了其他问题,这一模型可能产生带有交互效应【Interaction】但缺乏相应的主效应【Main Effect】的模型。于是,他们提议使用 Type Ⅱ分析(也被称作叶芝拟合常数法)。仅当其他因子不包含被估值的变量时,Type Ⅱ法才对其他因子予以调整。因此,变量 X_1 将由于 $X_1 \times X_2$ 的存在不予调整,因为后者包含前者。若一个模型仅包含主效应,Type Ⅱ分析则与 Type Ⅲ分析没有区别。在其他情形下,Type Ⅱ要求从待处理模型中选出应当剔除的因子。

若一个因子模型中,一个或多个格子是空的,得自 Type Ⅰ,Type Ⅱ或 Type Ⅲ分析的结果常无法解释。这时,可以使用技术上更复杂的第四种方法。Type Ⅳ SS 需要构建一个假想的、平衡的数据格矩阵,并以此为基础,来估计各自变量的效应。然而它的解可能不唯一,因此在解释其结果时需要格外谨慎。

——Douglas Madsen

(王玥译校)

＊也可参见方差分析模型 I【Model I ANOVA】。

参考文献

Nelder, J. A., & Lane, P. W. (1995). The computer analysis of factorial experiments: In memoriam—Frank Yates. *The American Statistician*, *49*, 382-385.

Yates, F. (1934). The analysis of multiple classifications with unequal numbers in the different classes. *Journal of the American Statistical Association*, *29*, 51-66.

无偏的(Unbiased)

当估计量【Estimator】$\hat{\beta}$ 的期望值等于参数【Parameter】的真值,即 β : $E(\hat{\beta})=\beta$ 时,估计值被称作是无偏的。就一个估计量而言,无偏性是一个期望的属性,因为它可以给予我们这样的信心,即一个参数的平均估计值等于真实的参数值。当参数的期望值不等于其真值时,即 $E(\hat{\beta})\neq\beta$ 时,这一估计是有偏的。偏倚【Bias】程度可以用估计的期望值减去参数的真值计算得出:偏倚 $=E(\hat{\beta})-\beta$。

思考无偏估计量的一种方式是通过想象两个人射箭和试图命中靶心。每个人是一个不同的估计量,而靶心代表被估参数的真值。个体 1 射出 100 支箭;箭支命中点的平均击中靶心。个体 2 也射出 100 支箭。然而,个体 2 击中目标的能力并没那么好,箭支略微偏左。个体 2 的箭支命中点的平均距靶心约有 1 码。箭支的平均落点可以代表由两个估计量产生的估计的期望值。个体 1 就是一个无偏的估计量,因为,平均而言,他或她能够命中靶心。个体 2 是一个有偏的估计量,因为箭支平均走向真值的左侧。

值得注意的是,无偏性是生成估计值的概率分布【Probability Distribution】的属性。一种无偏估计技术不会每次都给予我们待估参数的真值。毋宁,只有在重复抽样【Sampling】后,被估参数的平均值等于参数的真值。

不能因为估计量是无偏的就说它是理想的估计量(Wooldridge, 2002)。反而,估计量如果是一个无效的估计量(见效率【Efficiency】),这一无偏估计量并不能告诉我们有用的信息。一个无效的估计量的概率分布具有大的方差,从而导致任何一个给出的估计值等于参数的真值在概率上有高的不确定性。个别情况下,我们会使用有偏的估计量,因为它们有其他理想的属性。

无偏性是普通最小乘方【Ordinary Least Square, OLS】作为估计技术的一个有力特性。伍尔德里奇(Wooldridge, 2002)解释到,当四个条件得到满足时,OLS 是无偏的。第一,因变量和自变量间的关系在函数形式上是线性【Linear】的。第二,观测值抽取自一个随机样本【Random Sample】。第三,零条件均值的假定【Assumptions】。零条件均值陈述,给定 x,误差项的期望值为零。第四,被研变量上存在变化。当这些条件得到满足时,OLS 是一个无偏的估计量。当一个估计量表现出是有效的、有最小的方差且是无偏的时,它就是一类特殊的估计量,被称为最佳线性无偏估计量【Best Linear Unbiased Estimator, BLUE】。

——Heather L. Ondercin

(赵锋译校)

参考文献

Wooldridge, J. M. (2002). *Introductory econometrics: A modern approach*. Cincinnati, OH: South-Western College Publishing.

不确定性(Uncertainty)

社会科学家双重地关注不确定性。当人们试图建立关于一个不完全可预测的社会世界的模型时,不确定性就出现了,而形式研究方法则被设计来量化这种不确定性。但是,需要运用这些方法的人类行为中的大多数易变特征来源于个体社会成员根据他们自身的不确定性作出的判断。

经济学中的理性期待理论是上述两种考虑清楚相关的鲜有情形:股票的价格,这一个社会-水平的现象,被预知为随时间随机波动,因为所有有效的系统信息都已为个体投资者用来极小化关于股票价格的不确实性。任何剩余的价格变动反映的是其余投资者的不确定性。

概率【Probability】提供了相应的基础,以表述对不确定性的两种考量:大多数经济学和政治学理论假定,概率推断是一个关于在一个不确定的世界上个体决策制定的合理的行为模型。与特定社会科学理论有关的不确定性也应用概率和相关的统计学工具予以表述。但是概率不是量化不确定性的唯一选择,因此,值得思考的是,为什么这是个好的选择。

关于个人运用概率理论来管控不确定性的动机,有几种认证。"荷兰赌(Dutch Book)"论证说明在赌博的情境中,那些面对不确定性而以同他们可用的概率理论不一致的方式行事的个体总是被诈取(Ramsey, 1960)。求助概率理论的另一更一般的动机由理查德·考克斯(Cox,1946/1990)提出。考克斯说明,对于任何关于在一个可能构造的命题A上的确定性,似真或置信的数值测量,它都必须是这样的概率的简单再测,即如果它具有下列直觉的属性,A则是正确的:

1.传递性:如果A比B更似真,且B比C更似真,则A比C更似真。

2.A的似真度是非A的似真度的函数(例如,A越似真,则非A越不可能真)。

3.A和B同时的似真度是A的似真度与当A为真时B的似真度的函数。

如果这些属性应用于相应的量,那么我们分配给A的数值将是A的概率的常数倍。如果原则1~3被接受为最低的要求,那么,概率理论就是量化不确定性的正确方式。因为考克斯的论证不依赖被量化的不确定性是个人的抑或社会科学的,所以他的论证给用概率量化不确定性以一般的驱动。尽管在经济学中理性期望理论有广泛的应用,但是大多数心理学家怀疑概率理论是个体在不确定情况下推理的一个好的描述模型(Tversky & Kahneman, 1974)。例如,相对于既有的数据,人们过于自信,而且当他们做预测时,部分地忽略事件的先验概率(基本—比率忽略)。对理性期望理论而言,这些心理事实不是必然成问题的,因为个体期望只需要平均地同概率理论一致。它们也不会妨碍基于概率理论的个人推论理论;人们可能对于一个不正确的内在模型作出概率上正确的推论,或者考虑相应的因果限制(Cheng,1997)。人们似乎也以这样一种方式处理信息,即付出某些概率上正确推理【Inference】优势的损失,运用不正确,但是较少计算需要,仅仅接近正确的推理过

程而取得速度优势(Gigerenzer & Goldstein, 1996)。

科学的不确定性是否应当以个人的不确定性同样的方式予以量化的问题确立了统计学上频次论者和贝叶斯论者间的争论。我们用一个熟悉的例子的两个解释来讨论这一问题。

关于某个量 Y 和两个潜在的解释变量 X_1 和 X_2 间的关系的线性回归【Linear Regression】模型可以写作

$$Y = X_1 b_1 + X_2 b_2 + e$$
$$e \sim N(0,\sigma^2) \quad (1)$$

在该模型的贝叶斯分析中,参数值是不确定的且这一不确定性由先验概率【Prior Probability】分布 $p(b_1, b_2, \sigma^2)$ 予以描述。即使已知参数确定,结果 Y 也可能是不确定的。这一不确定性被分解成一个确定性的【Determinstic】部分,累加关系 $X_1b_1+X_2b_2$ 和一个随机的【Stochastic】部分,即假定随机变量【Random Variable】e 有一个特定的概率分布。关于 Y 的不确定性可通过观测 X_1 和 X_2 部分地减少,直到由 e 所确定的最小水平;e 可能代表一个真正随机的社会过程的效果或者我们的确信,其中存在着与 Y 相关但没有被包含进模型中的变量。方程(1)关于结果的概率 $p(Y|b_1, b_2, \sigma^2)$,即基于特定的参数值【Parameter】和解释变量(忽略已知的解释变量)的条件概率。贝叶斯定理【Bayes Theorem】结合这两个分布入一个后验概率分布 $p(b_1, b_2, \sigma^2 | Y)$。这一后验概率分布结合了参数值的不确定性和 Y 值的不确定性,而后者的不确定性始终存在,即便其他条件都已知。后验值的顶峰是参数最可能的取值,而一个 95%的区间具有 0.95 的包含真值的概率。

后验概率是基于 Y 的条件概率,因为在得到观测数据之后,它们的值已经确知。以观测数据为条件是唯一的可能,因为存在一个先验分布。对频次论者而言,应用先验分布于科学研究是不合适的,因为这会降低科学推论的客观性。在此情形下,概率被严格限于描述客观的随机机制,而不确定性被限于完全无知的状态。根据这一解释,e 描述了一个客观的、物理的或社会的随机过程,且参数固定,但完全未知。

为推论参数值,频次论者选择有效的 Y 值的函数作为估计量【Estimator】,再计算点估计【Point Estimate】和置信区间【Confidence Interval】。Y 由于包含 e 而有概率分布,所以参数估计(而不是参数自身)作为客观的随机变量 Y 的函数,也有其概率分布。置信区间通过仅仅间接地使用概率而反映不确定性:这是一个区间构造的方法,同一个概率(如 0.95)相联系,意味着在重复抽样中,以这种方式构造的 95%的区间将包含真实的参数值。因为观测到的 Y 值不可能作为条件,一个置信区间依赖于无数假设的结果——Y 值,这无法实际地出现,而是假设已经实现。

尽管根据重复样本的概率定义容许频次论者避免量化不确定的水平,但是这可能是有问题的。例如,许多社会科学研究关注于解释本质的唯一情境,其中重复样本的想法没有意义。相反,把 e 解释作一个不确实性的量度同任何关于重复抽样的可能性的观点是一致的,因为在两种情形下都存在着不确定性。

至于由参数的先验概率所表现的不确定性,重要的是看到后验分布表述,给定一组先验的信心,以观测数据为基础更新相应的信心是合理的。如果给出一组不同的先验值,那么合理的信心很可能不同。尽管先验概率不可避免地是主观的,但是它也有清晰的表达,且无歧义地被引入推论,同时,改变它的后果可能提供关于数据信心内容的有效信息。

频次论者和贝叶斯论者对社会科学推论的争论转到了新的方向,即概率是否是不确定性的最佳表现形式。换言之,以实际观

测数据为条件取得的知识增益(排除、抽样理论的局限,并认同概率是不确定性的量度这一基本论断)是否比下述命题更有价值,即只有将主观要素加进来,才能促进与社会世界有关的科学知识的进步。

——Will Lowe

(赵锋译校)

参考文献

Berger, J. O., & Wolpert, R. L. (1984). *The likelihood principle: A review, generalizations, and statistical implications.* Hayward, CA: Institute of Mathematical Statistics.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104* (2), 367-405.

Cox, R. T. (1990). Probability, frequency, and reasonable expectation. In G. Shafer & J. Pearl (Eds.), *Readings in uncertain reasoning* (pp. 353-365). New York: Morgan Kaufmann. (Original work published 1946)

Edwards, A. W. F. (1972). *Likelihood: An account of the statistical concept of likelihood and its application to scientific inference.* Cambridge, UK: Cambridge University Press.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650-669.

Ramsey, F. P. (1960). Truth and probability. In R. B. Braithwaite (Ed.), *The foundations of mathematics and other logical essays.* New York: Harcourt Brace.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, *185*, 1124-1131.

均匀关联(Uniform Association)

参见关联模型【Association Model】。

单峰分布(Unimodal Distribution)

这是一个单变量【Variable】的值的分布【Distribution】,仅有一个众数和单个"峰(peak)"。众数是三个用于确定一个变量的"典型分(typical score)"的常用量度之一,是一个给定变量最常出现的值(Lewis-Beck, 1995, p. 8)。

单峰分布的一个例子是标准正态分布【Normal Distribution】。该分布均值【Mean】为零,标准差【Standard Deviation】为1。在此例中,均值等于中位数【Median】和众数。而且,标准正态分布是唯一的分布,有相等的均值、中位数和众数。因此,它是一个单峰分布,因为它只有一个众数。

然而,一个单峰分布并不必须像标准正态分布一样有对称性【Symmetry】。一个单峰分布可以同时是偏斜【Skewed】的、左偏或右偏。此类分布的一个例子如美国的收入。

图1列出美国总体的、分类别的年度收入(DeNavas-Walt & Cleveland, 2002, p. 20)。收入的中位数,其中收入值的一半大于它,另一半小于它,落在类别$42 228上。这一数字说明美国家庭收入分布略微向左侧偏斜,因为中位数更接近于底部的10%(类别$10 913),而非顶部的10%(类别$150 499)。此外,众数类是$42 228。

与其他类别相比,有更大百分比的家庭(大约50%)在这一类别上,表示这个分布无疑是单峰的。

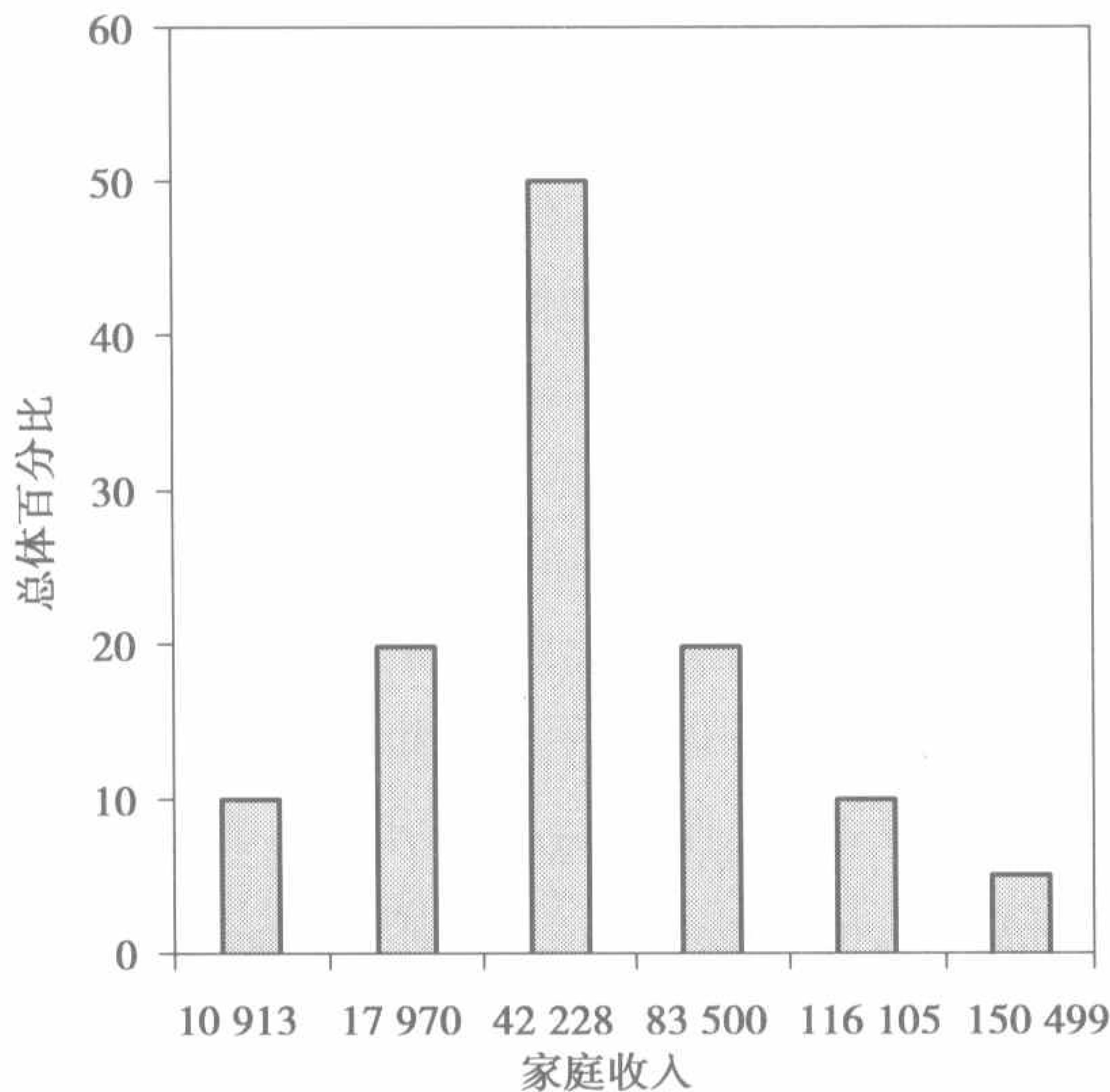


图1 位于不同被选分位数上的家庭收入直方图

无论一个单峰分布是偏斜的还是对称的,它必须只有一个"峰"。一个有单一"峰"的分布是这样一种,即对于任何给定的变量而言,只有一个最常出现的值,如它的频数分布【Frequency Distribution】所显示的那样。如果一个分布有两个"峰",那么它是一个双峰分布【Bimodal Distribution】。最后,如果一个分布超过两个"峰",它就是一个多峰分布(multimodal distribution)。

——Kenneth W. Moffett

(赵锋译校)

参考文献

DeNavas-Walt, C., & Cleveland, R. W. (2002). *Money income in the United States* (United States Census Bureau, Current Population Reports, P60-218). Washington, DC: Government Printing Office. Available: www. census. gov/prod/2002pubs/p60-218.pdf

Kennedy, P. (1998). *A guide to econometrics* (4th ed.). Cambridge: MIT Press.

Lewis-Beck, M. S. (1995). *Data analysis: An introduction.* Thousand Oaks, CA: Sage.

分析单位(Unit of Analysis)

分析单位是科学研究方案的最基本要素。它是研究的主题(谁或什么),而分析者要形成关于它的概念。在社会科学中,国家、国际联盟、学校、社区、利益群体和投票者是经济发展、战争、教育策略、社会资本、政策结果和投票研究中常用的分析单位。

分析单位可能同观测单位不一致。例如,政治科学家可通过估计关于政治参与测量的均值分,在国家的水平上聚合调查数据,并分配同样的分给相应的国家,以此来辨识跨国间参与模式的不同。在这种情况下,观测单位是个体的被调查者,而分析单位是国家。与此相似,学生的成绩也常被聚合起来,使其可以在学校间相比。这时,国家和学校是分析者所要进行概化的对象,而不是市民或学生(见聚集【Aggregation】中的慎重提示)。

在上述例子中,有过多的细节信息可用,甚至超过分析者的研究需要。然而,在社会科学中,这不是常例:常例是研究者常常检视未观测到的单位行为。例如,如阿尚和夏夫利那样,分析者可能希望考察个体的某个社会属性同投票行为间的关系,但是仅有地区水平上聚合的数据可用。也就是,他或她有的是居住在此地区,具有这一社会属性的人群的比例,以及支持该政党的比例。这种跨水平的推论,其中分析单位同观测单位不一致,会产生错误的结论,因此,应避免

这种推论(见区域谬误【Ecological Fallacy】)。

然而,即便分析单位同观测单位相匹配,社会单位的特征还是会令科学推论变得复杂。布克(Boucke,1923)提示了两点社会单位区别于自然科学中常用分析单位的差别:第一,社会科学研究单位不能直接测量,或是不能在一个确定的标尺上测量。常见的是,社会现象上相当比例的方差可以归之于测量误差(见系统误差【Systematic Error】)。此外,某些研究者有兴趣的单位只在它们的因果效应上是可观测的(见Blalock,1961)。第二,如前述例子所表明的,社会科学中的分析单位总是可以被分作多个子单位:社会可以被分作家或家庭、地理的邻里或者个体成员。布克论断社会科学中分析单位不是"事物(things)",而是将家庭、邻里或个体利益(最基础的水平上)联在一起的关系或网络。社会科学中的子单位鲜有如自然科学的子单位所有的实质意义或混合【Confounding】效应。此外,社会科学子单位间关系的变异度使得科学推论成为一项困难工作:社会研究中分析单位的混合性质产生"有风险的预测,而且越是这样,研究者就越过分关注心理的关系,而不是物质的关系"(Boucke, 1923, p.460)。社会科学的分析者需有效地开展工作,以便容纳他们所选分析单位的隐含意义于他们的研究设计。

——Karen J. Long

(赵锋译校)

参考文献

Achen, C., & Shively, W. P. (1995). *Cross-level inference*. Chicago: University of Chicago Press.

Blalock, H. M. (1961). Theory, measurement and replication in the social sciences. *American Journal of Sociology*, *66*(1), 342-347.

Boucke, O. F. (1923). The limits of social science: Ⅱ. *American Journal of Sociology*, *28* (4), 443-460.

单位根(Unit Root)

一个时间序列可以模型成自身过去的函数,即一个自回归。在它最简单的形式中,这样一个序列可写作 $y_t = \alpha + \beta y_{t-1} + \varepsilon_t$,其中所有 ε_t 项被假定互相独立。如果 $\beta = 1$,则该序列有单位根。单位根的存在对于解释和估计【Estimation】,无论是对简单的自回归,还是对带有一个滞后的因变量和其他自变量的更复杂的模型而言,都具有决定性的后果。关键是要检验一个自回归序列是否有单位根,如果有,还要在模型参数的估计中说明这点。

如果一个序列有单位根,那么可以将自回归形式重写为 $y_t = \alpha + \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \cdots$,因此,当前的 y 是所有此前误差的和[积分只是总和的一种形式,这也是何以此类序列被称作"可积的(integrated)"的原因]。因为误差项独立,所以 y 的方差随时间增大。因此,所有通常使得回归【Regression】的渐进理论有效的假定不再成立,而且所有标准的方法都不正确。特别地,如果自回归序列有单位根,但我们试着用普通最小二乘【Ordinary Least Squares】估计 β,那么我们将低估 β 且误估它的标准误。解释也受到极度的影响:若带有单位根,那么对于序列的小扰动将持续存在,而不是如其在固定序列(其中 $\beta<1$)上那样逐渐消失。对于可积的序列而言,中心极限定理【Central Limit

Theorem】也不再成立，而且在这种情形下，所有的标准分布理论也由此失去作用。

迪基和富勒是第一个为虚无假设（即一个时间序列有单位根）设计了正确检验的人。这一检验由两步构成，使用普通最小二乘估计 β，再形成 Dickey-Fuller 统计量$(1-b)/SE$，其中 b 是用普通最小二乘估计的 β，SE 是该估计的普通最小二乘标准误。若是固定时间序列，该统计量有 t 分布【t-Distribution】；在单位根的虚无假设下，迪基-富勒必须用艰深的数学来得出该统计量的适当分布。因此，使用者可以计算迪基-富勒统计量，再将其值与由迪基-富勒的制表分布相比较。现在，这一检验有许多现代的改进，但是它们都是基于迪基-富勒原来的想法。在应用 Dickey-Fuller 检验时，重要的是记住虚无假设是序列、是可积的。如果虚无假设被拒绝，标准方法就可以应用。若虚无假设未被拒绝，这也不意味序列有单位根——它只意味着我们没有充分的信息拒绝该假设。不幸的是，没有假设可以翻转上述逻辑，因此，如果 Dickey-Fuller 检验没能拒绝虚无假设，研究者通常假定序列有单位根。

如果序列有单位根，那么标准的普通最小二乘方法不再能应用。一个解决方案是取一阶差分并予以模型，因为一阶差分通常不会显示单位根。实质上，这仅仅等于模型一个序列的短期行为，而忽略该序列的任何长期行为。另一解决方案是面对单位根的分析者可以应用共积【Cointegration】方法论。

——Nathaniel Beck

（赵锋译校）

参考文献

Engle, R. F., Granger, C. W. J. (1991). *Long run relationships: Readings in cointegration.* New York: Oxford University Press.

Mills, T. C. (1990). *Time series techniques for economists.* Cambridge, UK: Cambridge University Press.

单变量分析（Univariate Analysis）

单变量分析是对单变量的统计特征的分析，包括它的分布【Distribution】、中心趋势【Central Tendency】和离散【Spread】。

单变量分析中，第一步是将观测值绘制成表格或图，进而检视变量的分布。定类【Nominal】变量可以用饼图【Pie Chart】或条形图【Bar Graph】呈现。在一个饼图中，饼中的每一角代表变量的一个类别。在一个条形图中，每个条的高度表示观测值在变量每个对应的类别上的数量或比例。定距【Interval】和定比【Ratio】数据常以直方图【Histogram】、盒形图【Boxplot】或分位数【Quantile】图来呈现。有些变量有对称【Symmetry】的分布，另一些变量可能有带异常值的偏斜的【Skewed】分布，而异常值【Outliers】会扭曲根据这些分布计算的统计量【Statistics】。

单变量的总结性统计量常被称为描述统计量【Descriptive Statistics】。集中趋势量综括了变量的典型值。算术平均称作均值【Mean】，常作为数值变量的集中趋势的量度。中位数【Median】表示中间位次观测值的变量值，用于定序【Ordinal】变量。众数【Mode】是最常出现的值，用于定类数据。当数据中存在异常值或变量的分布极端偏斜时，中位数常常用于代替数值数据的均值。

离中趋势【Dispersion】的量度表示一个变量的观测值中变化的量。就定类或定序数据而言,离中趋势的简单量度包括类别数和众数类别上观测值的数量。对数值变量来说,离中趋势的最简单量度是极差【Range】,即最大观测值同最小观测值的差。因为异常值会扭曲极差,所以一个替代量度是四分位距【Interquartile Range】,即第75个百分位数同第25个百分位数之差。在盒形图上,四分位差被明显地表示出来。就定距和定比数据而言,标准差【Standard Deviation】和方差【Variance】(标准差的平方)是离中趋势的通用量度。标准差和方差的较大值表示观测值间有更大的离中趋势。

我们还可以检视单个变量的分布,以确定其接近一个对称,钟形的正态分布【Normal Distribution】的程度。对称分布在中位数的两侧有相同的形状。偏斜度统计量表明一个分布背离对称的程度。峰态【Kurtosis】统计量表明多大程度上,一个分布围绕众数有高且瘦的峰形或是矮而平的峰形。

——David C. Kimball
Herbert F. Weisberg
(赵锋译校)

参考文献

Fox, W. (1998). *Social statistics* (3rd ed.). Bellevue, WA: MicroCase.

Jacoby, W. G. (1997). *Statistical graphics for univariate and bivariate data.* Thousand Oaks, CA: Sage.

Lewis-Beck, M. S. (1995). *Data analysis: An introduction.* Thousand Oaks, CA: Sage.

Weisberg, H. F. (1992). *Central tendency and variation.* Newbury Park, CA: Sage.

全体(Universe)

参见总体【Population】。

未观测到的异质性(Unobserved Heterogeneity)

未观测到的异质性是指观测值中潜在的或隐藏的特征。这些特征是没有测量到的或研究中未予测量的个体或子群体的属性。一般而言,个体或群体的特征会作为协变量【Covariate】(或自变量【Independent Variable】)纳入统计模型。当未测量到的特征影响被考察的过程时,我们说观测值中存在未观测到的异质性。

就某些统计模型而言,未能说明个体或群体间的异质性会导致有偏参数估计【Parameter Estimation】,结果的错误解释或虚假【Spurious】的时间-协变量交互。因此,明确地纳入未测量到的异质性,使其成为模型的组成部分的统计技术已经得到了发展。结果是模型中加入一个或多个参数,以描述在样本所取自的总体中,未测量到的特征是如何分布的。

例子

1985年孟加拉国农村地区的结婚年龄(见图1)。粗线表示不考虑性别(或未予测量)时,结婚年龄的分布。假设一个对数正

态分布被确认是结婚年龄的合适分布。拟合观测值到一个对数正态分布（不参照性别）产生模型参数，结婚年龄均值【Mean】为23.3岁（标准差【Standard Deviation】5.9）。未能考虑性别的因素会掩饰实质的子群体差异。实际上，一个设定性别的分析产生男性结婚年龄均值27.3（5.5），女性结婚年龄均值19.5（3.4）。子群体的差异非常大，以致对任何协变量（比如，教育对结婚年龄的影响）的参数估计都可能由于缺少性别的信息而是有偏的。

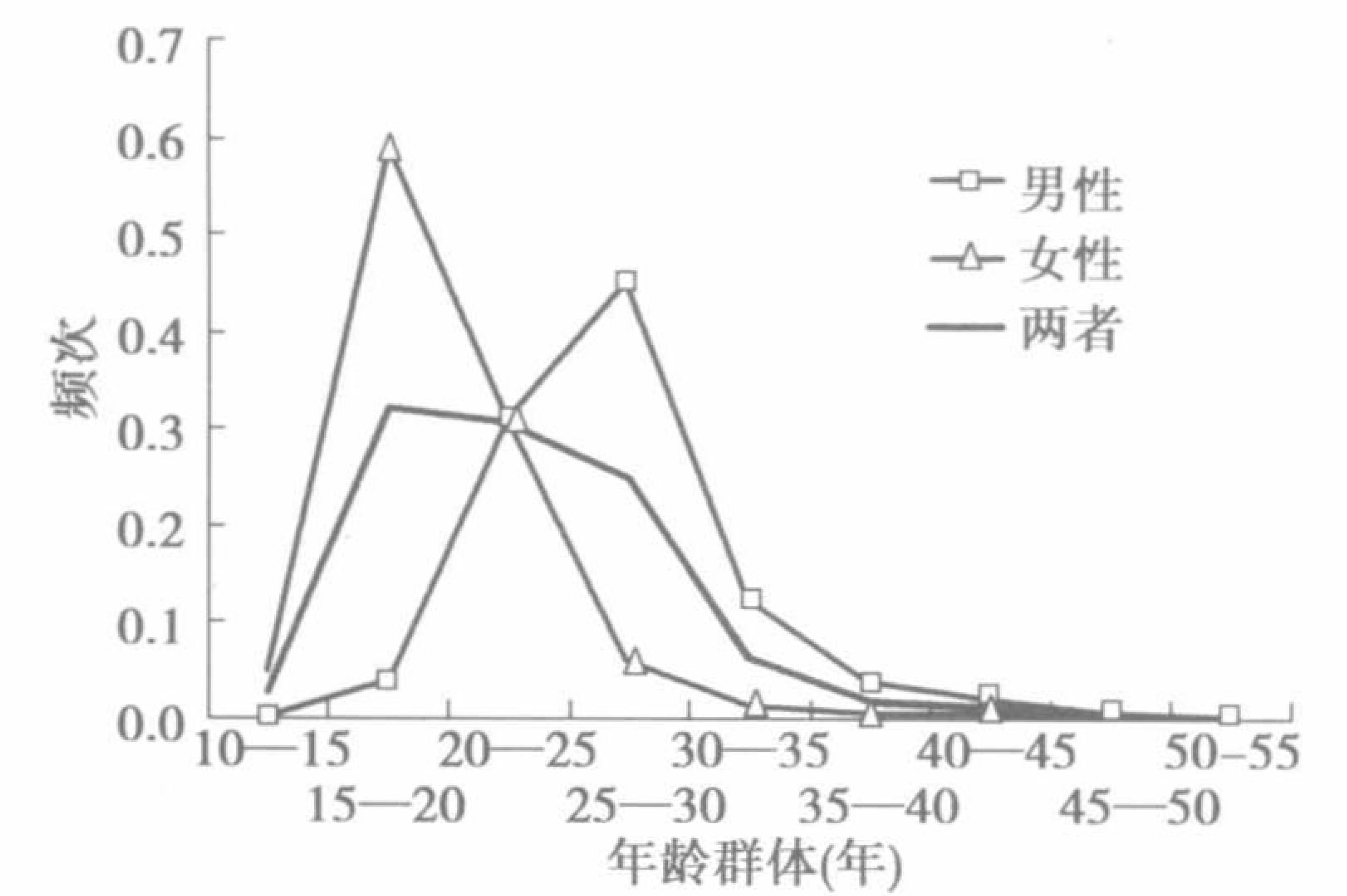


图1 1985年孟加拉国Matlab地区的结婚年龄

资料来源：国际腹泻疾病研究中心，孟加拉国（1992）。

在没有性别信息的情况下，一个解决方案是估计一个有限混合模型的参数。假定男性和女性数量相等，可以估计四个参数的混合模型，并得到一个子组的结婚年龄估计值为25.4（7.1），另一个子组的结婚年龄估计值为21.4（4.0）。两个子组各自对应于男性和女性。

风险模型中的“曲折动力学”

对生存分析和相关的方法而言——处理事件的持续时间（风险模型或事件史分析【Event History Analysis】），未能说明未观测到的异质性会导致严重的解释问题。就这些模型而言，未观测到的异质性是指观测单位中在它们“失败（failure）”（即经历所考察的事件的风险）的风险上存在的未测量到的差异。忽略失败风险上未测量到的异质性会导致风险随时间的下行偏估。

思考婴儿的死亡风险，最初这种风险非常高，过了第一年后，又迅速下降。就字面意义讲，这一观测意味着婴儿经历的是反向衰老：随时间推移个体“变得更好（get better）”。部分而言，这一模式可反映已知的生物过程，如免疫系统的成熟。另一方面，相同的“婴儿死亡”模型可以在其他复杂系统中看到，如计算机和汽车。人们不太可能真地认为个人计算机在早期服务期间经历了失败风险上的下降。表面的反向衰老是随时间变化的样本构成所产生的人为效应，“年龄同期群的曲折动力学”的产品（Vaupel & Yashin, 1985a, 1985b）。

图2显示出聚合的风险作为未观测到的异质性的结果，如何随时间而向下偏斜。两个子组出现在图上，一个位于高且恒定的失败风险上，另一个位于低且恒定的失败风险上。子群体最初含有相同数量的个体。当子群体被合并在一起且被当作一个单一同质的总体处理时，就会随时间变化观察到一个失败风险上的下降趋势。这一下降趋势之所以出现是因为属于高风险子群体的个体以更高的比率发生意外。而随着时间的展开，生存总体逐渐由低风险个体组成。

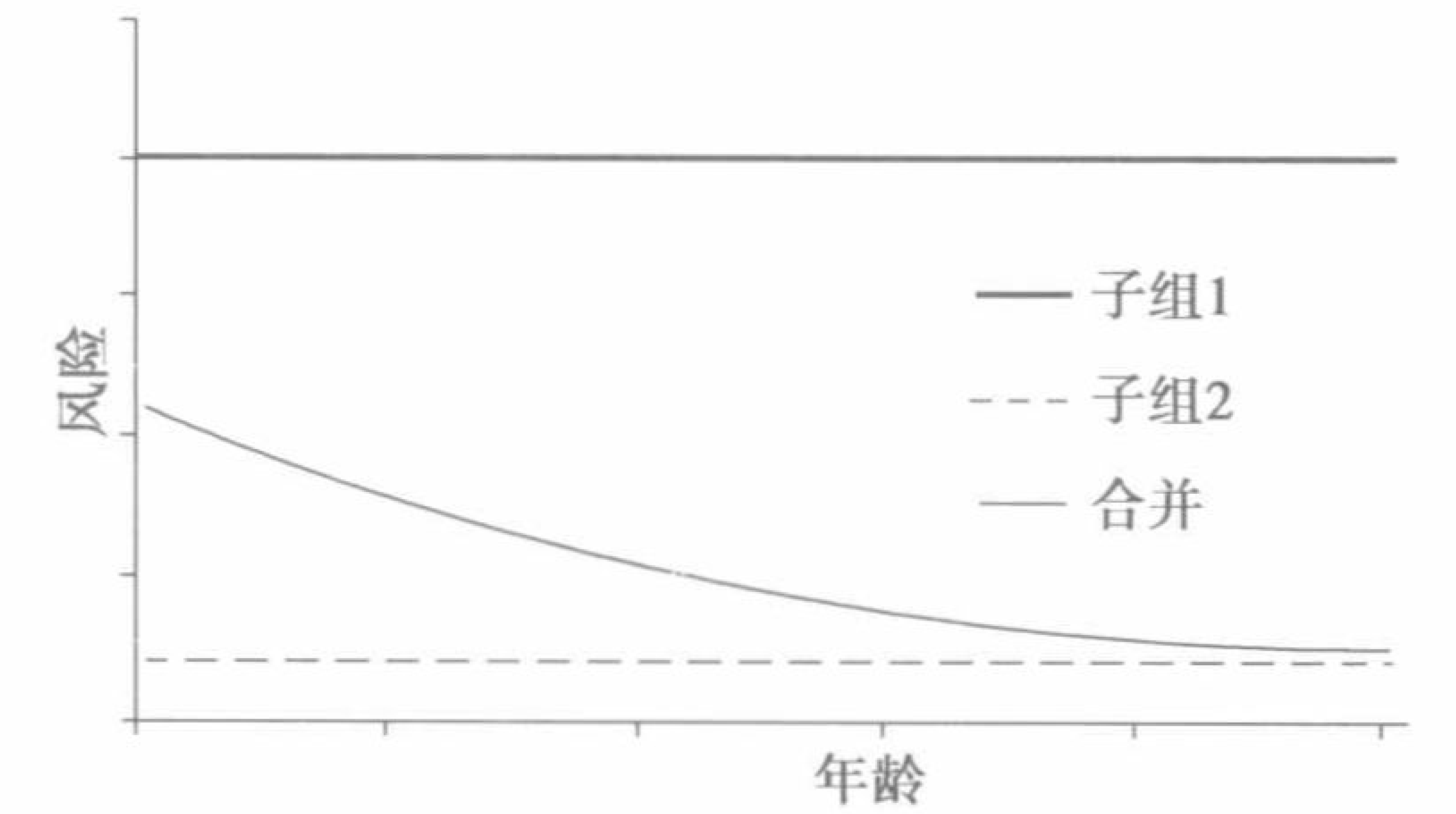


图2 两个同质子群体的混合，每个子群体有固定的失败（风险）率

注：当子群体被合并后，风险随年龄呈下降趋势

异质性的一个重要特例是潜在子群体之一的所有个体都不存在失败风险,例如,研究夫妻生育率的人口学者可能进行多次观测,从结婚直到每一对夫妇生育第一胎。其中有部分夫妇既不可能生育,也不会生育。而困难之处在于,某些能生育的夫妇,在第一个孩子出生前,就退出了研究或没能被持续观测。这些未能观测到他们第一次生育的夫妇来自两个根本不同的子群体:被右删的可生育个体和不能生育的个体。特定的分析模型被称为不育模型、免疫模型或移动者-驻留者模型(mover-stayer models),可用于估计第一胎之前等待时间的无偏分布,以及不育个体的初始比例。在许多社会科学研究中这类过程都会出现,包括结婚研究、离婚研究、生育跨度、复发倾向、职业流动和居住流动。

缺陷模型

在失败风险上未观测到的异质性可在个体间呈现一种接近一个连续分布的分布形态。一个重要的示例是由死亡率的人口研究所发展出的缺陷(frailty)概念(Vaupel, Manton, & Stallard, 1979)。个体会有无法观测的(或未观测到的)遗传的、环境的或行为的倾向或缺陷,而这些倾向或缺陷会以特有的方式调整他们的死亡率,使其偏离某个基线分布。某个个体缺陷可以由量 z_i 描述,但是因为 z_i 是不可测的,一个替代方案是估计参数的概率分布,而该分布描述了 z 是怎样在个体间分布的。

对许多类型的缺陷模型而言,z 对失败风险的效应可被设定成一个比例的风险模型,因此,对于时间 t 有 $h(t|z)=h(t)z$。选取的缺陷分布有均值 1,方差参数描述了缺陷在总体中分布的方式,也是待估的。最常用的参数型缺陷分布是 Gamma 分布。有时,当 z 的分布有均值零时,则使用模型设定$h(t|z)=h(t)e^{z}$;同时,常常应用正态分布【Normal Distribution】。

来自缺陷模型的结果可能对缺陷分布选择很敏感。另一种方法是将异质性的分布设定成一个配对的离散密度函数(Heckman & Singer, 1982)。众数点的数量和位置是由经验确定的。当理论无法建议一个合适的 z 分布时,这种方法可能特别适用。即使如此,这一策略可能对用于风险函数的参数模型是敏感的(参见 Trussell & Rodríguez, 1990)。

除了缺陷模型,还有许多其他可用的未观测到的异质性的参数模型。例如,二项分布【Binomial】和几何【Geometric】分布可用作所考察事件发生率(p)的异质性的分布。当且仅当分布位于0~1 时,常用一个 Beta 分布作为 p 的分布。

——Darry J. Holman

(赵锋译校)

*也可参见异质性【Heterogeneity】。

参考文献

Heckman, J., & Singer, B. (1982). Population heterogeneity in demographic models. In K. Land & A. Rogers (Eds.), *Multidimensional mathematical demography* (pp. 567-599). New York: Academic Press.

Heckman, J., & Singer, B. (1994). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Ecomometrica*, *52*(2), 271-320.

International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDRB). (1992). *Demographic surveillance system—Matlab: Registration of demographic events—1985* (Scientific Report 68). Dhaka, Bangladesh: Author.

Manton, K. G., Singer, B., & Woodbury, M. A. (1992). Some issues in the quantitative characterization of heterogeneous populations. In J. Trussell, R. Hankinson, & J. Tilton (Eds.), *Demographic application of event history analysis* (pp.

9-37). Oxford, UK: Clarendon.

Moreno, L. (1994). Frailty selection in bivariate survival models: A cautionary note. *Mathematical Population Studies*, *4*, 225-233.

Trussell, J., & Rodríguez, G. (1990). Heterogeneity in demographic research. In J. Adams, D. A. Lam, A. I. Hermalin, & P. E. Smouse (Eds.), *Convergent issues in genetics and demography* (pp. 111-132). Oxford, UK: Oxford University Press.

Vaupel, J. W. (1990). Relatives' risks: Frailty models of life history data. *Theoretical Population Biology*, *37*, 220-234.

Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, *16*, 439-454.

Vaupel, J. W., & Yashin, A. I. (1985a). The deviant dynamics of death in heterogenous populations. In N. B. Tuma (Ed.), *Sociological methodology* (pp. 179-211). San Francisco: Jossey-Bass.

Vaupel, J. W., & Yashin, A. I. (1985b). Heterogeneity's ruses: Some surprising effects of selection on population dynamics. *American Statistician*, *39*, 176-185.

Vermunt, J. K. (1997). *Log-linear models for event histories.* Thousand Oaks, CA: Sage.

Wood, J. W., Holman, D. J., Yashin, A., Peterson, R. J., Weinstein, M., & Chang, M.-C. (1994). A multistate model of fecundability and sterility. *Demography*, *31*, 403-426.

Wood, J. W., & Weinstein, M. (1990). Heterogeneity in fecundability: The effect of fetal loss. In J. Adams, D. A. Lam, A. I. Hermalin, & P. E. Smouse (Eds.), *Convergent issues in genetics and demography* (pp. 171-188). Oxford, UK: Oxford University Press.

非强行测量(Unobtrusive Measures)

非强行测量是韦布、坎贝尔、施瓦茨和西克里斯特对非强行研究程序的经典参考书的书名(1966;也参见 Webb, Campbell, Schwartz, Sechrest, & Grove, 1981 年的修订版)。李(Lee,2000)也出版了一本极好的概论;也请参见非强行方法【Unobtrusive Methods】。替换了测量一词,非强行方法,这涵盖更广的术语,成为当前更常用的术语,因为它包含了除韦布等人专注的定量方法之外的其他方法。

"非强行测量"是一套研究程序,其设计目的在于避免许多社会科学研究中的一个主要问题,即反应性【Reactivity】问题(Webb et al., 1966)。研究参与者一般对他们的情境作出反应;此即,他们觉察或怀疑到他们正在以某种方式被观测或评估。如果一个人用尺子测量长度,测量过程不会影响观测的精度或效度。然而,如果使用问卷、抽样调查或访谈程序评价对一个问题的态度时,数据很可能是"不纯的(impure)",也就是说,受参与者的反应和特质的影响。例如,关于种族包容或性问题的问卷【Questionnaire】或调查可能引起参与者对于研究目的的反应性顾虑(例如,谁会看到这些数据及什么样的回答会被视为适当的)。数据对研究人员的某些因素也是敏感的,例如性别、种族、年龄或其他特征。而且,有些研究参与者还想"帮助(help)"研究者,而另一些有时则不太合作,甚至极不合作。尽管有时他们的假设不太准确,但是参与者常依照这些猜测进行回答。

另一种常见的情况是,研究者自己也想或希望他们的研究能有圆满的结果。我们知道这些期望可能影响研究如何开展(Rosenthal, 1976),以及后续的结果。

与之不同的是,非强行方法是那些参与

者没有觉察到他们正在被观察或被用作数据源的方法。这里,研究者确信重要的是要避免如下常规要求,即获得参与者的事前同意,也就是说,他们的积极合作。非强行研究经常或者彻底地避免人的卷入,而使用历史的或档案的资料,或者使用如隐藏录音机、麦克风,甚至隐藏的观察者等研究策略。其他的研究则以研究“说和做(say and do)”为特征,此即,他们将某些情境下参与者会说的或会做的同他们实际被观测到的行为进行对比。比如,一个人可以打电话给那些刊登了出租房间或公寓广告的房东,提出一些简单的求租征询,然后将征询结果同基于其他条件的征询结果——诱导房东相信致电者属于某一人种或具有某种可耻的特征,如是同性恋者、有艾滋病或曾经是精神病患——相比较。李(Lee,2000)认为可以把互联网作为非强行研究的领域,如对主页(home page)的分析。

因为非强行测量的数据没有获得参与者的积极合作或知情同意,所以这些方法不可避免地遭遇伦理困难。然而,社会科学的一个主要目标是收集、比较和对比观测数据,而无论其数据源自强行的还是非强行的测量。

——Stewart Page

(王玥译　赵锋校)

参考文献

Lee, R. (2000). *Unobtrusive methods in social research*. Buckingham, UK: Open University Press.

Rosenthal, R. (1976). *Experimenter effects in behavioral research* (Rev. ed.). New York: Irvington.

Webb, E., Campbell, D., Schwartz, R., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences*. Chicago: Rand McNally.

Webb, E., Campbell, D., Schwarz, R., Sechrest, L, & Grove, J. (1981). *Nonreactive measures in the social sciences*. Dallas: Houghton-Mifflin.

非介入方法(Unobtrusive Methods)

多年来,社会科学家一直强调研究的形式方面——因此,我们有许多关于统计学和研究设计【Research Design】的教科书。但有一个遗留的问题:我们对于研究应该如何开展的认识比对于它实际的开展要全面得多。在心理学中,许多研究假定,如人们会直接写下或口头告诉我们他们的所思、所感或将要做的,或者假定他们只会对研究的“刺激”自动地作出“反应”。然而,我们经常发现,人们所说的(或他们对他们作为的述说)与他们实际做的形成鲜明对照。一个对照是,当他们知道他们正在被观察时,他们的行事方式与当他们假定他们的行为是私人的、不被知晓的、不被他人观察时的所做所为。非强行方法(unobtrusive methods)是一组研究方法的名称,而这组研究方法既不依赖也不寻求研究的人类参与者的积极卷入或知情同意,还常常以参与者不察觉的方式开展研究。一些方法,例如内容分析【Content Analysis】,从非人来源中收集数据和观察资料,如杂志、图书、报纸、网络材料、媒体演示或历史文献。

背景

一个使用非强行方法观察人类行为的例子是美国电视系列片《偷拍》(*Candid Camera*),它虽然恶名昭彰,却仍为社会科学家所重视。这个系列片的基石是对那些处

于各种真实的社会情境,又不知晓自己被观察的人的非强行摄录。这些被摄录的人径直地"做自己(be themselves)"——而后常常陷入尴尬。例如,*Candid Camera* 的职员,在大山顶上把引擎从汽车中搬出。然后汽车在没有动力的情况下被"开(driven)"下山,直接滑到山下的服务站。当(女)司机说她觉得引擎"抛锚(running rough)"了,几名(男)服务站工作人员打开车盖仔细地查看,然后显出一副被他们发现的引擎问题难倒的样子,即便实际上并无可看的引擎。

非强行(unobtrusive)指无干扰、无声息或在背景中。因此,非强行方法指出的是一个简单却经常被遗忘的教训。此即,只要情况许可,我们就应考虑使用那些不掺杂自身要素,或者不以某种方式影响或改变待收集数据的性质的研究方法。然而,并不存在可以唯一界定非强行研究方法的特定类型的定量或定性数据分析。而且,对于如何界定它们,既不存在限定条件,也不存在严格的界限。非强行一词仅指一个一般的方法种类。

非强行方法的历史发展源自1960年代以来一个不断增强的认识,即多种形式的社会研究,特别是那些涉及研究者与研究参与者间的社会互动的研究,常常受到许多不可靠和偏差来源的影响。这些方法的共同目标是要避免韦布、坎贝尔、施瓦茨和西克里斯特(Webb, Campbell, Schwartz, & Sechrest, 1966)提出的反应性【Reactivity】,即许多传统(强行)方法趋向于向参与者泄露或与参与者沟通研究的内容,比如,研究者正在寻找什么。然后,参与者再作出"反应(react)";即他们以不同于下述条件的方式行事——设若他们的资料通过非强行的方式或没有察觉到自己是某个研究的焦点。反应性的一个形式是参与者在数据收集时根据相应的线索,觉察或正确地猜出研究的目标或假设的内容。而且,我们知道,有许多参与者或者以制造出"好印象(good impression)"的方式,或者根据其他形式的偏差,来回答问卷【Questionnaire】或调查【Survey】中的题项。另一些情况下,参与者对研究内容的猜测可能不确定或不准确,但仍会影响他们的回答和行为。事实上,存在着数据反应性的多重来源,不仅有参与者对于研究的性质和目的的知觉和假设,还包括由研究者和研究情境本身的特征造成的影响。

韦布等人(1966; 也参见 Webb, Campbell, Schwartz, Sechrest, & Grove, 1981 的修订版)就此类问题的认识发表了《非强行测量》(*Unobtrusive Measures*)一书。社会科学领域从这些作者那里受益匪浅,因为我们对于这些程序的知识和意识大部分都可以追溯到他们的著作。李(Lee, 2000)就这一领域给出了令人赞许的概述和许多最近的例子。

韦布等人强调他们并不完全反对传统的研究方法。这点却常被忽略。不过,他们指出,这些测量有时是必需的,经常以轻信且独自的方式为人所用,此即,没有用可以考察同一研究问题或议题的非强行方法做补充或去验证。基于此种情况,社会科学就难以建立其研究发现的效度界限和概推界限。我们也很少知道,比如,传统的检验或数据收集程序能多精确地预测真实生活(不同于实验的或实验室的)情境中的行为。前面提到的传统方法为人所熟悉的有:访谈、心理测验、定性研究、等级量表、抽样调查,以及许多一次性的,基于简单的——常常非常明显或透明的——问卷测量的研究。因此,在社会科学中,我们经常发现,显微镜的存在和特征实际上会造成所研究材料或有机体的行为的改变。

非介入方法的例子

至此,我们可以把非强行或非反应方法描述为试图避免反应性,且不要求或寻求所

涉参与者的积极合作、意识或知情同意的一类方法。而许多这样的研究被称为"说和做(say and do)"研究，或更准确地，"说对做"研究。其实，李(Lee，2000)把这类研究划归为田野实验【Field Experiment】。下面列举了几个例子，相关的参考文献和进一步的细节可参见佩奇(Page，2000)的文章：

1.通过非强行的测量，而后比较不同展品的周围楼层表面，每单位时间内表面损坏的程度来衡量图书馆的资料或博物馆的藏品的流行度。

2.根据瞳孔扩大的程度表明顾客对购买特定商品的兴趣度。

3.通过观测正在做广告出租房间或公寓的房东就某些个体能否租住给予虚假信息的程度，估计一个给定地区的种族偏见度。在观测中，使得房东相信真与之沟通的个体属于某个特定种族或具有某种可耻的特征，如是同性恋者、艾滋病患或曾经是精神病患。这一基本方法的不同变体已被用于争取公民权和废除种族隔离的努力中，此即，在调查中，由黑人作为潜在租户去寻求那些已知没有租出且可租住的房间，有时要亲自拜访户主家收集这一信息。

4.通过分析非强行获取的(在参与者不知情的情况下)录音记录，如在线的、电话的或现场的互动的录音，估计临床治疗师或悲痛中心咨询师表现出他们对案主的同情程度，以及相关的帮助技能水平。

5.在何种程度上，那些实际上对黑人低接纳和低容忍的回答人会表现出对黑人的高接纳和高容忍。例如，通过访谈、心理测验或调查问卷确定回答人报告的接纳度和容忍度，再经过一段时间后，通过非强行的观测，即在自助餐厅，他们更多地光顾白人收银员而非黑人收银员，或者在服务过程中，显出细微的(如非言语的或姿态上的)拒斥或敌意，估计他们实际的接纳度和容忍度。

6.在何种程度上，相比于没有患过精神疾病的人，潜在的社区雇主会把岗位给予或不给予那些宣称曾经患有精神疾病的人。康涅狄格大学的阿梅里戈·法里纳教授和同事已用这种方法做了几项研究，有时还使用了隐藏的录音设备。

7.在何种程度上，具信是写给某个精神病患，而被意外"遗失(lost)"在各种地点的信会被其发现者转给(寄给)收件人。

8.在何种程度上，可以发现学生对其课程或教授的两种评估方式间的差异，一种基于访谈或强制性的书面评估，一种来自对学生在等候室或自助餐厅的对话的非强制记录。

9.在何种程度上，通过非强制观察可以发现，黑人顾客可能在商店受到的服务比白人糟糕，或者在购买旧车或其他商品时，相比于白人，被给予更高的报价。

10.在何种程度上，一段时间后，大学生的品行和校风的变化，可以反映在大学书店里所售服饰或其他商品上印制的大学的校色、校标或校名的变化上。

11.可以依据"精神疾病(mental illness)"这个标签，检查报纸上的文章和新闻的内容，来判断一个社会对待精神疾病的态度的性质以及一段时间内的变化。人们可以如此开展此项研究，例如，在加拿大报纸索引中，依照此标题(精神疾病)，查看所有存档的加拿大报纸的新闻和故事，然后比较其近期和前期的内容。使用存档的或"持续的(running)"记录(Lee，2000)或内容分析【Content Analysis】(例如，从即存资料中获取观测数据)——而不是直接引入或观察参与者——是当下一种流行的非强行数据收集方式。

12.最后一个例子涉及使用吸引力不断增强的互联网去收集非强行获取的数据和观察资料。互联网呈现了一种数据资源，并且容许以非强行研究的方式探查在线用户的自我表达的内容和方法。另外，这一领域的研究者极大地摆脱了通常研究在时间、空

间和费用上的限制。李(Lee,2000)的文献是对互联网为基础的(internet-based)研究的极佳讨论,列述了互联网研究的诸要点,还给出了许多例子。例如,对主页(home page)风格和内容的研究可以成为了解个人、群体或组织的一种手段。有时,关于敏感主题的研究信息可以通过建立可靠的和保密的在线联系来收集。许多互联网行为(behavior)的匿名性让我们有机会可以研究那些处于虚拟环境中,完全不在乎他人如何"看(look)"的在线用户。同样,可以研究计算机的使用和沟通模式,以及聊天室(chat rooms)中的在线发帖模式、在线讨论模式或新闻组模式。

伦理考量和非介入方法的现状

此外,欲评估非强行方法的现状还要考虑获得参与者的知情同意【Informed Consent】的问题。首先,我们注意到许多在韦布等人(Webb, et al.,1966)著作第一本书的时代流行的通用研究方法,现在仍然很受欢迎并被普遍使用。这一点能反映出对同意问题的普遍重视;此即,它们之所以仍然盛行,不是因为它是效度的固有的或绝对的保证,而是因为通常研究者获取和记录参与者愿意参加的同意书并不困难。这一观点与现在流行的社会建构论(social constructionism)是一致的,即研究参与者必须被看成真正参与到其提供数据的研究之中。由此,建构主义观点把参与者看成帮助研究者创造结果,好像他们同研究者之间具有的研究关系(research relationship),是一种共存类型。基于此种观点,事实和知识并不是被发现的(像我们一般假定的那样),而是被发明的或是由社会共识创造出来的。(心理学的研究参与者直到最近,被称为受试者,这一词对于社会建构主义者来说,意味着他们是被动的,像物体一样,而不是帮助研究者创造知识的人。)

于是,非强行方法由于通常在没有研究参与者的合作或知晓的情况下进行资料收集,所以很难在研究的方法论或认识论上取得当下的盛行。当然,对于韦布等人(Webb, et al.,1966)来说,尽管在某些背景下,参与者的积极参与、知情同意及合作仍然是适用的,但是它们不是研究的根本目标。相反,它们代表了要去避免的反应性的来源或至少需要由其他测量来补充。大体上,非强行测量的真正目的和目标在于有效地拒绝对特定类型数据的依赖和寻求——此种数据需要积极获得研究参与者知情同意(和合作)(Page, 2000)。尽管我们关注反应性方法的概推性和效度,但是在当下学生的研究课程或研究工作中,非强行方法很少获得强调。据韦布等人在他们先前著作的修订版(Webb, et al.,1981)中所作报告及李(Lee,2000)的转引,尽管大部分的社会科学家看似都承认非强行测量的价值,但目前这种方法的实际应用并不常见。

非强行测量确实涉及伦理陷阱,同时,在一些情形下,它们的使用甚至在法律上是不妥当的。在参与者不知情的情况下,记录或观测特定行为,或者如一些已经实施的研究,在公共场所创造模拟的突发情境,并观测公众的反应,这些做法可能是有问题的或令人难堪的。有时,从非强行方法得到的数据可能并不比由其他方法得到的数据更可靠或更容易解释,因此,在某些场合,人们必须慎重地考虑它们。另一问题是,如果替代的方法并没有受到反应性的过度影响,且可以获得参与者的知情同意,那么,一项研究是否应避免非强行测量。李(Lee,2000)还描述了几项由于使用互联网进行非强行研究带来的伦理问题。这些问题都涉及在公共的和私人的领域间划定界限的困难,以及保护在线参与者的秘密,而这点至为关键。

这里,我们提请读者注意,韦布等人(Webb, et al.,1966)没有把非反应性方法视作天然的非伦理的,实际上,把它们看成可以让研究者高枕无忧的方法,即在不引发参与者生活上的不便、担忧或搅扰的情况下收集资料。同韦布等人(Webb, et al., 1966, 1981)和其他人的著作相一致,关于非强行方法不断增多的考量认为,这种方法具有补充传统方法并引导我们走出对反应性数据依赖的潜力,从而有效地让我们再次认识到研究观测的概推性这一基本问题以及实施三角测量【Triangulation】的必要性——使用不同类型的测量来研究某个特定的问题或假设。我们还提请读者注意,唐纳德·坎贝尔确信社会科学研究最终应该为了他人而开展,即出于共同体和社会的利益。应用某种方式的非反应方法的研究,虽然有时还涉及具有一定科学合理性的对参与者的某种欺骗,但是具有生成惊人的、意料外的,常常也是反自觉的结果的潜力——也因此而具有特别的兴味和价值。

——Stewart Page

(王玥译　赵锋校)

参考文献

Lee, R. (2000). *Unobtrusive methods in social research.* Buckingham, UK: Open University Press.

Page, S. (2000). The lost art of unobtrusive measures. *Journal of Applied Social Psychology*, *30* (10), 2126-2128.

Webb, E., Campbell, D., Schwartz, R., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences.* Chicago: Rand-McNally.

Webb, E., Campbell, D., Schwartz, R., Sechrest, L., & Grove, J. (1981). *Nonreactive measures in the social sciences.* Dallas: Houghton-Mifflin.

非标准化的(Unstandardized)

非标准化的概念可以同术语标准化的相对照而获得最佳理解。标准化是将不同公制的统计量转换到一个共同的公制上,从而允许比较。换言之,如果统计量是非标准化的(有不同的测量条件),它们就不能相比。以多元回归方程的非标准化的偏回归系数【Partial Regression Coefficient】或斜率为例。考察这样一个有关巴西的时间序列数据,1970—1990(World Bank, 1997):$\hat{Y}=-5\ 764+110X_1+28X_2$,其中 $\hat{Y}$ 是以美元计算的人均国民生产总值(Gross National Product, GNP),X_1 是城市人口的百分比,而 X_2 是以10亿美元为单位的资本净流入。两个变量,X_1 和 X_2,检验都显著。结果表示巴西城市人口百分比增加1%,平均而言,伴随有110美元人均GNP的增加,而增加10亿美元流入巴西的净资本则伴随有28美元人均GNP的增加。然而,非标准化的偏回归系数或斜率,并不说明哪个变量,城市人口还是资本净流入,对因变量,即人均GNP有更大的影响。尽管,就城市人口百分比而言,其斜率值更大,清楚地表明城市人口增加1%完全不同于资本净流入增加10亿美元。事实上,可以将资本净流入的测量单位从10亿美元变为兆美元,进而人为地增大资本净流入的斜率(非标准化的偏回归系数)。斜率的大小,即非标准化的回归系数,并不意味那个自变量同因变量上的更大变化相关联。

为便于比较自变量【Independent Variable】对因变量【Dependent Variable】的影响,我们可以在一个共同的标尺上将变量标准化,从而产生标准化回归系数【Standardized Regression Coefficient】,也称之为Beta权。计算自变量和因变量距离各

自均值相对于标准差【Standard Deviation】的位置,可以将它们标准化。这一程序会产生新的变量值,再用这些新的值重新估计回归方程(在分析数据时,该程序可以由统计软件实现)。Beta 权或标准化的回归系数是“保持其他自变量不变的情况下,伴随着变量 X 上一个标准差的变化,变量 Y 上平均的标准差的变化”(Lewis-Beck, 1980, p. 65)。这一转换现在容许估计哪一个自变量同 Y 上的更大变化相联系。在该例中,我们发现城市人口上 1 个标准差的变化伴随着人均 GNP 上 0.99 个标准差的变化,相比之下,资本净流入上 1 个标准差的变化只在人均 GNP 上伴随有 0.17 个标准差的变化。无疑,城市人口对于解释人均 GNP 的变化似乎更重要,资本净流入则不太重要。

标准化程序利用了正态分布【Normal Distribution】的已知属性,从而容许比较一个模型的自变量。然而,偏回归系数的标准化仅仅容许同一个样本内的比较。Beta 权不能跨样本比较,例如,确定城市人口百分比在阿根廷是否伴随有相似大的人均 GNP 的变化,假定巴西和阿根廷之间自变量和因变量的方差【Variance】不同。因此,跨样本的 Beta 权的比较可能人为地改变 Beta 权的值,进而可能导致错误的结论。相反,如果在两个样本中有相同的公制,用非标准化的回归系数,跨样本地比较一个变量,如城市人口对人均 GNP 的效应是适当的,因为非标准化系数对不同的 X 的方差不敏感。

标准化的概念还用在其他统计量度和概念上,如标准正态分布、标准分【Standard Score】和相关系数【Correlation Coefficient】。相关系数是两个变量间关联【Association】强度的量度。这一量度不会随测量单位的变化而改变,而测量单位的改变会影响协方差【Covariance】的量,因为相关系数只是通过使用每个变量的标准差计算其标准分,再得到标准化的协方差。协方差的标准化给出了一个关联的统计量,它更符合直觉,比起协方差统计量也更有用。因为协方差统计量缺少上界和下界,所以它不能提供关联强度的量度。我们不知道,协方差值的巨大是否反映了一个强的关联,抑或只是变量的测量单位的结果。标准化通过迫使关联的量度位于-1 和 1 的区间内,以-1 或 1 表示完全的线性关联,从而取消了上述问题。

标准正态分布和标准分为许多统计技术的实施提供了基础。这里,关于 Beta 权和相关系数的例子同时使用了这两个概念。标准分(也称作 Z 分数或标准化的变量)可以通过使用标准正态分布的概念,标准化相关变量而获得。某个变量 X 的标准分可以表示成变量值 x_i 减去样本均值,再除以标准差 s,其结果是一个带有标准正态分布(均值为零,方差为 1)的变量。如上述例子所示,标准化的本质是发现一个共同的公制,以此为基础,有不同公制的统计量或变量可以被评估。

——Jacque L. Amoureux

(赵锋译校)

* 也可参见标准化回归系数【Standardized Regression Coefficient】。

参考文献

Lewis-Beck, M. S. (1980). *Applied regression: An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-022). Beverly Hills, CA: Sage.

Lewis-Beck, M. S. (1995). *Data analysis: An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-103). Thousand Oaks, CA: Sage.

World Bank, International Economics Department. (1997). *World tables of economic and social indicators, 1950-1992* [Computer file]. Ann Arbor, MI: Interuniversity Consortium for Political and Social Science Research.

非结构访谈(Unstructured Interview)

非结构访谈包括不同的类型:创造性的、积极的、后现代的和女性主义访谈。他们享有共同的关注,允许被访者尽可能自由地回答开放式问题【Open-Ended Question】,并在自己的指导下回答。访谈被视作在访谈者与被访者间完成的一项互动。

非结构访谈,也被称为深度访谈,并不依赖于封闭式【Closed-Ended】或结构化的问题。访谈者通过提开放性的问题或只是提示被访者,求取某个既定主题的信息。访谈可能被录音,或者在访谈期间,极少在访谈刚结束后,访谈者做访谈记录。

为了访谈取得成功,访谈者必须获得回答者的信任。访谈者可通过自己的学术地位或研究领域获得信任,或通过暂时成为被观察群体的一员而获得信任;有时,可以付费给回答者让他们参与访谈(比照 Malinowski, 1989)。非结构访谈常与参与观察【Participant Observation】一起开展;实际上,芝加哥学派对 1930 年代的许多经典研究几乎完全依赖于非结构访谈(Harvey, 1987)。

近年来,非结构访谈出现了许多互相有别的形式。它们可以分为以下子类:

创造性访谈:在创造性访谈中,根据他自己以及他的研究生的研究,杰克·道格拉斯(Douglas,1985)提出了非结构访谈的不同技术。传统上,访谈者试图保持中立,不以任何方式影响研究。因此,作为个体,访谈者是谁同访谈是没有关系的。道格拉斯感觉不是这样的。他论证说,在非结构访谈中,一项无根据的假设是,回答者将如实地去回答。然而,道格拉斯扪心自问,如果人类在日常生活中总是撒谎,那么为什么回答者就有所不同呢? 为了实现最大限度的真实和全面公开,道格拉斯认为访谈者要给予补偿。访谈者必须向回答者揭示自己私生活中的一点隐私,从而真正地成为一个人,而不是一个没有人情味的、中立的访谈者,这样才能希望获得回答者的信任。此外,道格拉斯指出,通过回答者获取的信息必须通过使用其他知情人不断地检查其真实性。知情人最好是所研究群体的边缘成员,并因此心怀不满,才可能去泄露他们群体的信息。

积极访谈【Active Interview】。詹姆斯·霍尔斯坦和贾布尔·古布里乌姆(Holstein & Gubrium,1995)不再把访谈当作单向的事务——访谈者利用技术,最大限度地从被访者那里收集最大量的信息。代之,霍尔斯坦和古布里乌姆把非结构访谈看成访谈者与回答者间的积极互动。所得资料远不是中性的。毋宁,访谈是两个个体或更多个体共同达成的协商结果。该结果依赖于双方(两方共同创造了一个故事——访谈)合作的水平和互动的能力。访谈不发生在真空中,而是受它发生的背景的影响。对霍尔斯坦和古布里乌姆而言,访谈的另外的相关维度取决于,结果是如何达成的(对影响访谈的要素的反思性分析),以及达成了何种(what)结果(访谈的实质发现)。

后现代访谈。后现代访谈是非结构访谈的另一个子类。后现代主义【Postmodernism】是一系列渗透在许多学科中的哲学观念。尽管大可争议,但是它已然给非结构访谈带来了一些重要的思考。后现代主义拥护元理论和范式【Paradigm】之终结,希望应该转向对日常生活的细枝末节的研究和理解——像社会成员自身那样去看。显然,由后现代的观念,非结构访谈可以获得新的理解,特别是在表述对于非结构访谈的权力结构的关注时。传统上,访谈者控制访谈——提出问题、决定什么是相关的、确定访谈持续的时间,以及选择如何使

用访谈结果。后现代主义者认为,回答者被利用了,而且以无利于他们的目的去用他们和他们的信息是不道德的。后现代社会学家拒绝此种访谈模式,而偏向于公开的政治合作式的访谈。他们提倡对无特权群体的研究,辅之以帮助他们的开放议程。这是对于中立和客观性【Objectivity】的传统社会学观念的直接抗议。因为访谈结果将以帮助受研究群体的方式而被应用,所得结果就是公开的派性偏见(partisanship)。

女性主义访谈。最后一个子类来自女性主义访谈。通常,女性主义访谈联系着长期的口述史【Oral History】方法。女性主义者把传统的非结构访谈看作高度家长作风的(paternalistic)和有偏见的。传统访谈中,访谈者和研究的关注点均倾向于完全地以男性为中心。对于女性主义者而言,应该留意受访女性的关注点。同时,通过引入女性访谈者,访谈中的紧张得到极大缓解。女性对女性的访谈打开了一扇门,带来更加开放的信息流露,因为被访者摆脱了他们在传统的男—女访谈中的从属地位。于是,她们可以无所顾忌地讲出那些先前难以启齿的问题和议题。女性主义的访谈目的在于帮助被研究的女性,而不是利用她们和她们的信息。

总之,记住上述所有的变种,非结构访谈是探索社会学知识的基础工具。提出问题和获取答案的方法论允许社会学家去理解日常生活对其成员的意义。访谈的无结构特质并不排除可能的回答,却保留了一个完全开放的领域,从而培育对社会更广泛的理解。

——Andrea Fontana

(王玥译　赵锋校)

参考文献

Douglas, J. D. (1985). *Creative interviewing*. Beverly Hills, CA: Sage.

Harvey, L. (1987). *Myths of the Chicago school of sociology*. Aldershot, England: Avebury.

Holstein, J., & Gubrium, J. (1995). *The active interview*. Thousand Oaks, CA: Sage.

Malinoswki, B. (1989). *A diary in the strict sense of the term*. Stanford, CA: Stanford University Press.

专注于效用的评估(Utilization-Focused Evaluation)

专注于效用评估的前提是,评估【Evaluation】应根据效用和实际应用来评判;因此,评估者应通过从头至尾、仔细地检讨政策实施过程中的每一件事将怎样影响其应用,来促进评估过程和进行评估设计。关注点在于目标用户的预期应用。评估者需要发展同目标用户的关系,以帮助他们确定需要哪种评估。专注于效用的评估不建议任何特定的评估内容、模式、方法、理论,甚或应用。毋宁,它是一个针对主要目标用户的特殊处境,帮助他们去选择最适当的内容、模型、方法、理论和应用的过程。评估者同确认的主要用户群合作,共同努力于对他们的预期应用的评估。

——Michael Quinn Patton

(王玥译　赵锋校)

参考文献

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: Sage.

V

V

参见对称量度【Symmetric Measures】。

效度(Validity)

社会科学研究中最重要的步骤之一是对理论构念【Construct】进行测量【Measure】。为了能够经验地验证和检验研究中的现象,将理论【Theory】所描述的抽象概念【Concept】同那些概念的经验指标【Indicator】联系在一起是极重要的。一个抽象的理论概念同它的经验显现的关联是效度的核心。

一般而言,效度被定义为,测量工具在多大程度上测量了它所要测量的内容。然而,严格地讲,一位研究者能够估计一次测量的效度的唯一方式即通过验定"得自于一个具体程序的数据的一项解释"的有效性(Cronbach, 1971, p. 447)。此处清楚地隐含着,一个测量工具可能对测量某个现象是相对有效的,但是对于测量另一现象则可能是无效的。因此,测量工具本身是不用效度确认的,但是要确认测量工具同它所用于的目的的关系的效度。

在此我们要考察三种基本效度类型:效标效度【Criterion-Related Validity】、内容效度和构念效度【Construct Validity】。每一种类型以不同的方式评估效度,同时每种都有其用途和局限。

效标效度或称为预测效度是最接近上文效度定义的一种。如果一个量具对于某些其他的、外在的测量或准则有精确的预测,该量具即被认为相对有效。例如,通过表明一项书面的航空飞行员考试对于一群人能实际驾驶飞机的水平有精确的预测,人们可以确认该考试的有效性。效标效度的程度由测验和效标之间的对应程度来表示。事实上,这种相关【Correlation】是唯一一种与效标效度相联系的证据。

效标效度有两种具体的类别。一方面,如果效标当下已经存在,并且与一个当前的测量相关,那么我们即在运用并行效度(concurrent validity);另一方面,预测效度聚焦于一种情形,即测量预测了一个未来的效标或结果。于是,这两类效标效度的差异在于考虑效标的时间位置——当下或是将来。

内容效度是另一种基本效度类型。这类效度专注于一项具体的经验测量多大程度上反映了一个专门领域的内容。例如,一项对世界历史的测验如果只关注了美国或者欧洲的历史,那么它就不具备内容效度。

在评估内容效度时,一个研究者必须考虑以下几个方面。首先,研究者必须能够指明相关内容的整个领域。在上文的例子中,一项对世界历史的测验需要涵盖世界的所有地区,以及有记载的整个历史时期。对于社会科学研究中发现的许多抽象理论概念而言,通常难以指明内容的整个领域。其次,研究者必须取得测验领域的一个适当的样本。这里的难点在于,并不总是能够从整个内容领域得到样本。最后,关于内容的样本必须以适宜测验的形式呈现。因此,虽然直觉上内容效度具有吸引力,但是在社会科学中内容效度通常难以估计。

效度的第三种基本类型是构念效度。从根本上说,构念效度关注于根据理论期望,一个量具表现得有多好。评估一个量具的构念效度包括三个特定但又相互关联的步骤。首先,我们必须指明两个或更多理论构念之间的理论关系。其次,必须考察构念的量具之间的经验关系。最后,经验关系的证据必须依据它如何厘清特定量具的构念效度而予以解释。因此,如果基于理论思考,可以预见种族歧视在受教育程度较低的人群中的普遍程度比在受教育程度高的人群中高,人们可以通过检查一个种族论的测量是否同教育负相关,来评估它的构念效度。这个例子表明,构念效度只能在一个给定的理论背景内进行评估,因为这一生成相关理论预测的背景才是此类效度的核心。因为构念效度评估被直接地编织在社会科学理论的网中,所以,在社会科学中,它被视为比效标效度或内容效度都更符合效度的概念。

——Edward G. Carmines

Jame Woods

(马妍译　赵锋校)

参考文献

Carmines, E. G., & McIver, J. P. (1981). *Reliability and validity assessment* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-017). Newbury Park, CA: Sage.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Spector, P. E. (1992). *Summated rating scale construction: An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-082). Newbury Park, CA: Sage.

变量(Variable)

一个变量是其值会变化的事物,同常量——总是有同一值——相反(如数字2)。变量是事物可观测的特征,可以取几个不同的值,或者可以被置入几个不同的类别。例如,学生在某项测验上的得分就是变量,因为它们有不同的取值,同时,宗教也可以被当作变量,因为可以有多个种类。科学家有时对确定的常量感兴趣,如 π,圆周的面积同其半径平方之比。然而,统计学主要研究变量而非常量。

如果,对于任意数 a,当 X 小于或等于 a 时,存在一个概率 p,则量 X 是一个随机变量。一个离散随机变量是只取得若干确定值的随机变量,例如,某次选举中新候选人的数量。相反,连续随机变量则是可以在某个范围内取任意值的随机变量,例如,一个人的身高[用1英寸(1英寸=2.54厘米)的最小可能的分数予以测量]。

数据分析常常包括关于变量间关系的假设【Hypothesis】,例如“如果 X 的值增大,

那么 Y 的值也趋于增大（或减少）”。如果假设涉及潜变量间的关系，即抽象概念间的关系。这些概念还必须化成显变量，以便它们在实际研究中可以被测量。

在统计分析中一个基本的区分是因变量【Dependent Variable】和自变量【Independent Variable】间的区分，前者是研究者欲以解释的，而后者则作为因变量的预示变量。例如，在回归【Regression】中，因变量是 Y 变量【Y Variable】，位于回归方程 $Y=a+bX$ 的左侧，而 X 是自变量，位于方程的右侧。

统计分析的起点常常是逐一检视感兴趣变量的分布，包括计算适当的单变量统计量。变量随时间的变化还可用时间序列分析【Time-series Analysis】予以检视。而单变量分析通常只是双变量分析【Bivariate Analysis】或多元分析【Multivariate Analysis】的起点。例如，在方差分析【Analysis of Variance】中，研究者考察在多大程度上实验条件会影响因变量的方差。

——Herbert F. Weisberg

（赵锋译校）

参考文献

Lewis-Beck, M. S. (1995). *Data analysis: An introduction*. Thousand Oaks, CA: Sage.

变量参数模型（Variable Parameter Models）

人们通常假定在联系自变量【Independent Variable】和因变量【Dependent Variable】的回归【Regression】方程中，参数 beta（β）对于所有的观测值都是恒定的。但是对于时间序列【Time Series】数据，人们可能将参数 β 模型成随时间变化的，并假定这一变化遵从某些共同且连续的时间序列过程。因此，最简单的时间序列变量参数模型可写作，$Y_t=\beta_t X_t+\varepsilon_t$，其中 $\beta_t=\rho\beta_{t-1}+v_t$。

这样的模型可以由最大似然估计【Maximum Likelihood Estimation】法经过 Kalman 过滤器（Kalman filter）来估计。人们也可以检验参数是否随时间变化（或者如上面的模型呈连续的变化，或者以更加离散的方式变化）。这些模型也被称作随机参数回归模型（stochastic parameter regression model）。

——Michael S. Lewis-Beck

（赵锋译校）

＊也可参见随机系数模型【Random-Coefficient Model】。

参考文献

Newbold, P., & Bos, T. (1985). *Stochastic parameter regression models* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 51-001). Beverly Hills, CA: Sage.

方差（Variance）

参见离差（散布）【Dispersion】。

方差成分模型(Variance Components Models)

统计模型,如回归【Regression】或方差分析【Analysis of Variance】将观测到的变异分解成一个同解释变量或实验处理相连的系统部分和一个随机部分。通常,随机变异由单个误差项代表。当两个或多个随机变异的来源影响某个现象,变异的随机部分可以用一个方差成分模型来表示。

方差成分模型在社会科学中的应用包括,对来自多阶抽样【Multistage Sampling】设计的数据分析,用以估计测量质量的概化研究(见概化理论【Generalizability Theory】),包括随机因子的实验以及带有重复测量的纵贯研究。方差成分模型构成了多层次分析【Multilevel Analysis】或层级线性模型【Hierarchical Linear Model】的随机部分。

考察一下这样一个简单的模型,观测值为 y_{ij},来自两阶段样本:$y_{ij}=\mu_y+a_i+e_{ij}$;μ_y 是总体均值,a_i 是第一阶段的误差项,而 e_{ij}是第二阶段的误差项。第一阶段的误差项 a_i 被称作随机效应。在教育研究中,一阶抽样单位可能是学校,第二阶段的阶抽样单位就是校内的学生,而量 y_{ij}表示第 i 个学校中被抽中的第 j 个学生的数学成绩。这时,一种来源的随机变异(学生)内嵌于另一种(学校)之中。对应的方差成分模型是 $\mathrm{Var}(y_{ij})=\sigma_a^2+\sigma_e^2$,其中 σ_a^2 是来自第一阶段抽样单位(学校)的方差成分,而 σ_e^2 是第二阶段抽样单位(校内学生)的方差成分。

方差成分测量了 y_{ij}中可归之于不同误差来源的随机变异组成。例如,量

$$\rho=\frac{\sigma_a^2}{\sigma_a^2+\sigma_e^2}$$

给出了总随机变异中属于第一阶段单位(学校)的比例。ρ 称作组内相关【Intraclass Correlation】,也测量了聚集于第一阶段单位内的第二阶段单位间同质的程度。

在做统计推论时,应用方差成分模型可以将观测值间的相互依赖纳入解释。在上述例子中,人们可以从每个 M 学校的 n_i 个学生所构成样本,或者根据总的观测值 $N=\sum_{i=1}^{M}n_i$,来估计 μ_y。某个学校内学生的 n_i 个观测值并不独立,因为它们共享一个学校水平的误差值 a_i。在分析样本时,若假设它含有 N 个独立的观测值,会导致对 μ_y 的估计有太小的标准误;即 μ_y 的信度区间太窄,同时,与 μ_y 的假设检验相连的 p 值也会太低。

方差成分的估计通常假设随机效应呈均值为 0 的独立且相同的分布,并据此进行。方差分析(ANOVA)法利用如下事实,即在方差分析中期望均值的平方可以表示作方程成分的线性组合,从而将方差成分作为观测所得的均值平方的函数来估计。最大似然【Maximum Likelihood,ML】法和受限的最大似然(REML)法通过取得某个似然函数的最大值,获得方差成分的估值。这就要求关于随机效应分布形式的假设,而通常都设定其为正态分布。ML 和 REML 已经成为估计方差成分的首选。而这些估计值通常由迭代技术获得。

作为一个例子,上述模型可以应用到 1993 年一般社会调查的样本上,其中包含对来自 111 个初始抽样单位(PSU)的 1 015 名回答者的 10 项词汇测验。REML 参数估计值分别是 $\hat{\mu}_y=6.00$,$\hat{\sigma}_a^2=0.44$ 和 $\hat{\sigma}_e^2=4.09$。总方差为 4.53,而被估的组间相关系数是 0.10:大约 10%的词汇分的方差存在于 PSU 之间,而剩余的 90% 的方差在 PSU 之内。$\hat{\mu}_y$ 的估计标准误是 0.09,因此,均值 95%的置信区间是(5.82, 6.18)。如果假定1 015

个观测值是独立的,从而忽略回答者在 PSU 内的聚集再分析数据,均值的标准误为 0.07,而 μ_y 的置信区间也相应地变得更窄。

研究若多于两个随机变异来源,就需要更复杂的方差成分模型。根据随机变异来源的数量,以及它们相互交叉或嵌套的方式,这些模型还将采取不同的形式。例如,在一项关于工作表现评定的概化研究中,几个评分者可能根据若干项工作内容来对几个雇员的工作质量分别打分。评分者、工作内容和雇员在评定过程中是随机变异的交叉来源。若假定三个来源间不存在交互效应,对于评分 y_{ijk},即由第 j 个评分者对第 i 个雇员在第 k 项工作上打分的模型是 $y_{ijk}=\mu_y+a_i+b_j+c_k+e_{ijk}$,其中 μ_y 是评分的均值,而 a_i,b_j 和 c_k 分别是雇员、评分者和工作内容的随机效应。方差 $\mathrm{Var}(y_{ijk})$ 模型作 $\sigma_a^2+\sigma_b^2+\sigma_c^2+\sigma_e^2$,其中四个方差成分分别指雇员、评分者、工作内容和误差。这些方差成分可用于计算概化系数(相似于组间相关系数),以此来反映评分过程的可靠程度。

在多元回归模型中,参数中的随机变异会导向其他的方差成分模型。一个例子是在第 i 个学校的样本中第 j 个学生的数学成绩(y_{ij})对其 IQ(x_{ij})的简单线性回归:$y_{ij}=a_i+b_ix_{ij}+e_{ij}$。具体学校的截距 a_i 围绕某个平均的截距随机地变化,如 $a_i=a+\mu_i$,而具体学校的斜率 b_i 则围绕某个平均的斜率 b 随机变化,如 $b_i=b+v_i$。对于学校水平的回归参数的随机效应 u_i 和 v_i 是得分 y_{ij}中的随机变量源,因此,回归方程可以重写作 $y_{ij}=a+bx_{ij}+u_i+v_ix_{ij}+e_{ij}$。由此,给定 IQ 下成绩的方差,即 $\mathrm{Var}(y_{ij}\mid x_{ij})=\sigma_u^2+x_{ij}^2\sigma_v^2+2x_{ij}\sigma_{uv}+\sigma_e^2$,其中,$\sigma_u^2$ 是学校水平对截距的方差成分,σ_v^2 是学校水平对斜率的方差成分,σ_e^2 是来源于学校内学生的方差成分,而 σ_{uv}是学校水平对截距和斜率的随机效应的协方差。在带有多个嵌套水平和多个随机系数的模型中,可能出现相当复杂的随机变异结构。

——Peter V. Marsden

(赵锋译校)

参考文献

Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components*. New York: Wiley.

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

方差膨胀因子(Variance Inflation Factors)

方差膨胀因子(VIFs)测量了回归【Regression】分析中预测变量间的多重共线性【Collinearity】对估计【Estimation】精度的影响。问题中的方差是回归系数【Regression Coefficient】的抽样【Sampling】方差。方差膨胀因子这一术语可能由唐纳·马夸特(Marquardt,1970)造出。

思考一下线性回归模型

$$y_i=\alpha+\beta_1x_{1i}+\beta_2x_{2i}+\cdots+\beta_kx_{ki}+\varepsilon_i$$

其中,y_i 是 n 个观测值中第 i 个因变量的值;x_{1i}, x_{2i}, …, x_{ki} 是预测变量【Predictor Variable】;ε_i 是误差项,通常假定不同观测值的误差项相互独立且服从均值零,恒定方差 σ^2 的分布;α 是回归常量;而(在预测变量是数量型的模型中)$\beta(s)$ 是各个斜率系数。

β_j 的最小二乘【Least-Squares】估计值 b_j 的抽样方差是

$$V(b_j) = \left[\frac{\sigma^2}{(n-1)s_j^2}\right]\left[\frac{1}{1-R_j^2}\right] \quad (1)$$

其中，$s_j^2 = \sum(x_{ji} - \bar{x})^2/(n-1)$ 是第 j 个预测变量的方差，而 R_j^2 是 x_j 对其他 x 回归所得的多元相关的平方。方程(1)右侧的第二个因子 $1/(1-R_j^2)$ 被称作 b_j 的方差膨胀因子或 VIF_j，表示多大程度上预测变量间的共线性降低了 b_j 的估计精度，即相对于有类似散布，但不相关的预测变量而言。方差膨胀因子的平方根表示共线性对 β_j 的置信区间的影响大小。

这里需要提醒的是，由方程(1)还包括其他的、更普遍存在的、估计的不精确的来源：大的误差项方差 σ^2；小的样本量 n；以及带有小离散值 s_j^2 的预测变量。这里还需指出共线性必须非常强，才能导致回归估计精度实质地降低。图 1 画出了方差膨胀因子的平方根对多元相关 R_j 的曲线图；如图 1 所示，即便 R_j 大到 0.8，VIF 的平方根仍然小于 2。而预测变量间的多元相关要大到0.8，在社会科学数据中是鲜见的。

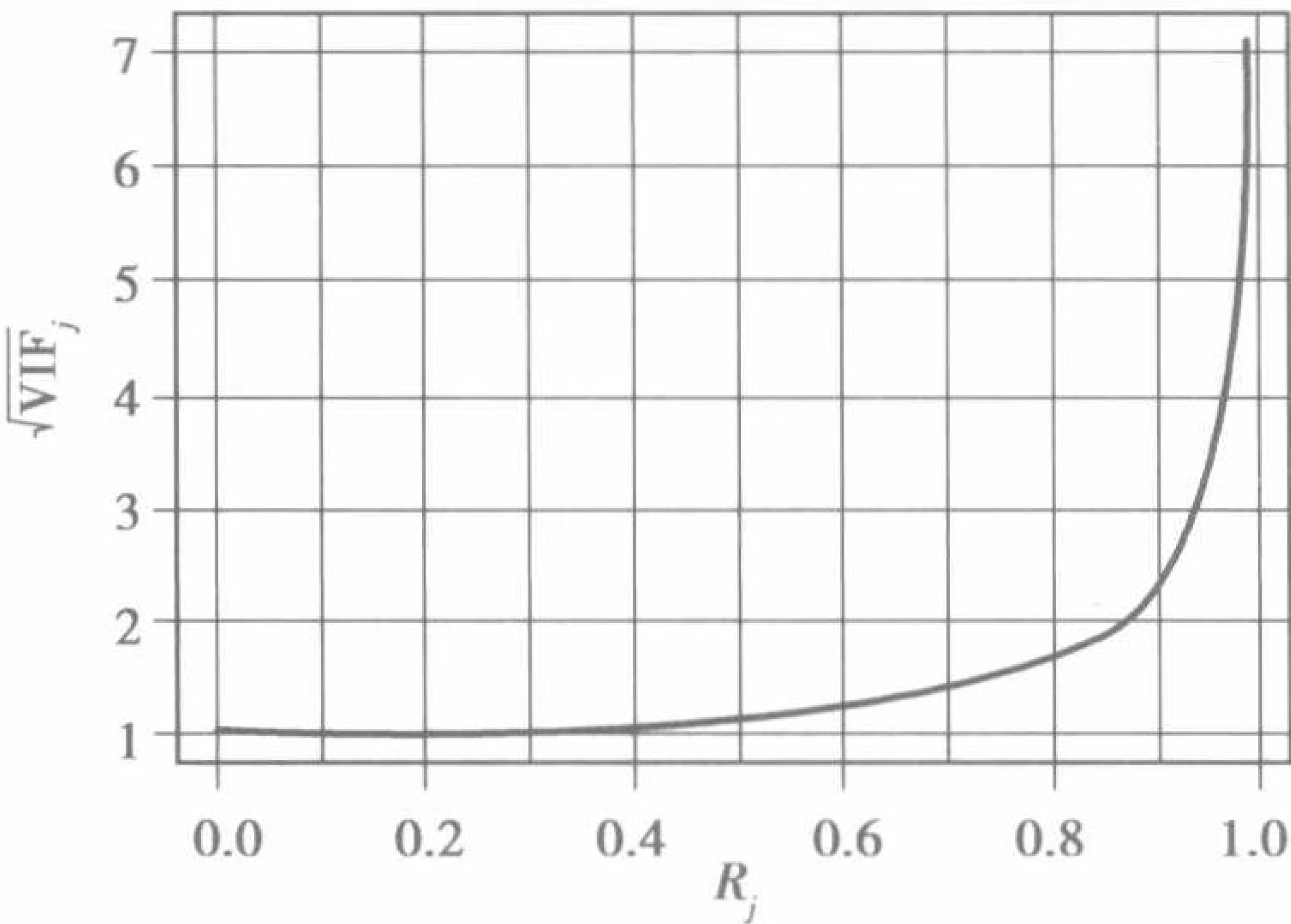


图 1　方差膨胀因子 VIF_j 的平方根，作为来自变量 x_j 对其他预测变量回归的多元系数 R_j 的函数

方差膨胀因子适用于线性模型中的单一系数(one-coefficient)项，如带线性效应的数量型预测变量。多系数(multiple-coefficient)项，如用以代表某个类别预测变量的一组二分回归项——需要更一般的方法，因为一组关联的回归项间的相关受到对模型无关紧要的变化的影响(如对一组二分回归项而言，改变其基准类别)。约翰·福克斯和乔治·莫内特(Fox & Monette，1992)通过比较两个回归系数子集(相关的预测变量和不相关的预测变量)的联合置信区间的大小，为此类情形提供了方差膨胀因子的一个一般解。

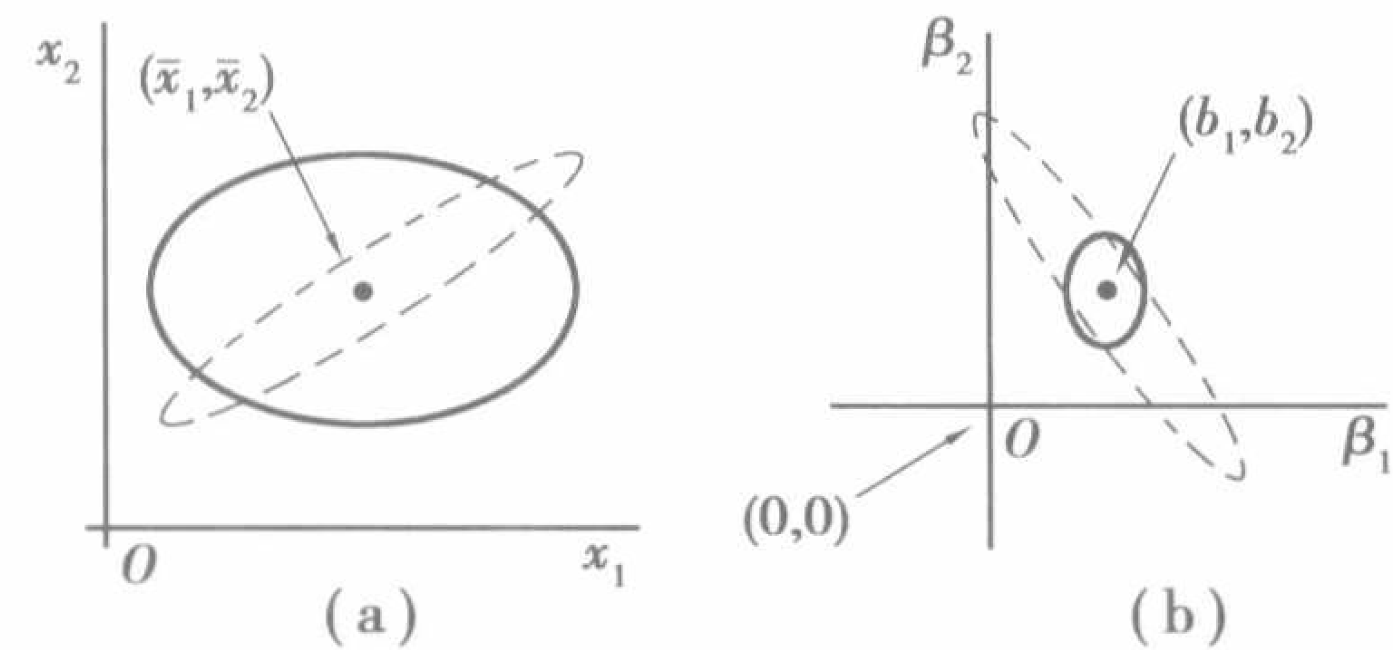


图 2　(a)无关(实线)预测项及高度相关(虚线)预测项的数据椭圆；(b)对应的 95%的置信椭圆

两个回归系数的联合置信区间由一个椭圆所限定；对于三个系数，它是椭圆体的，而对于三个以上的系数，它就是超椭圆体的。图 2(a)给出两个"数据椭圆"：实线椭圆用以表示不相关的预测项 x_1 和 x_2；椭圆的两个轴则分别平行于数轴 x_1 和数轴 x_2。虚线椭圆用以表示高度相关的两个预测项，其 $r_{12} = 0.95$；椭圆的倾斜反映了两个预测项的正相关。两个椭圆都是标量的，因此，它们在数轴 x_1 和 x_2 之上的投影是预测项的标准差；同时注意，两个椭圆都以预测项的均值为中心。就两变量的正态预测项而言，数据椭圆也是一个恒定的概率密度等高线，但是无论 X 的联合分布是什么形式的，数据椭圆是预测项的散布和相关的绘图呈现。

图 2(b)中的椭圆给出了回归参数 β_1 和 β_2 的 95%的置信区间——假定等于样本回归系数 b_1 和 b_2，这些置信椭圆是对应数据椭圆 90°旋转后的再标量。注意，高度相

关的预测项的置信椭圆远大于无关预测项的椭圆,同时前一个椭圆对于 β_1 和 β_2 而言,各自含有0值(尽管不必同时)。

现在,思考一下,把模型中的回归项分成两个子集:子集 $\mathbf{x}_1$,其系数是所有感兴趣的集合,补充的子集 $\mathbf{x}_2$,包含剩余的回归项。一般方差膨胀因子(Generalized Variance Inflation Factor,GVIF)比较了同子集 $\mathbf{x}_1$ 相连的联合置信区大小的平方和同子集 $\mathbf{x}_2$ 无关的、类似的、剩余的、回归项的联合置信区大小的平方:

$$\mathrm{GVIF}_1 = \frac{\det\mathbf{R}_{11}\det\mathbf{R}_{22}}{\det \mathbf{R}}$$

其中 $\mathbf{R}_{11}$是子集 $\mathbf{x}_1$ 中的回归项间的相关矩阵;$\mathbf{R}_{22}$是子集 $\mathbf{x}_2$ 中回归项间的相关矩阵;$\mathbf{R}$ 是所有 x 的相关矩阵;而 det 是矩阵行列式。当只有一个回归项在子集 $\mathbf{x}_1$ 中,GVIF 退化为通常的 VIF。

——John Fox

(赵锋译校)

参考文献

Marquardt, D. W. (1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, *12*, 591-612.

Fox, J., & Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association*, *87*, 178-183.

变差(Variation)

术语变差指一个变量【Variable】的离差或散布性【Spread】。如果在这个变量上所有的观测值只有同一个值,就不存在变差。当观测值都不相同时,即有变差。变差可被视作该变量异质性的程度。

如果不存在变差,就不必有统计学这一领域。人们感兴趣的变量是有变差的那一类,而统计学则试图估计变差的来源。中心趋势【Central Tendency】量显示,就一个变量而言,哪个值是典型的,而变差量则用以显示该值的典型程度。

就数值数据而言,变差定义作观测值偏离均值【Mean】的平方之和(Blalock, 1979, p. 338)。如果 X_i 是变量 X 上的第 i 个观测值,且设 $\overline{X}$ 是均值(见 X-条形图),那么,变差 $= \sum(X_i - \overline{X})^2$。一变量的方差【Variance】则定义为该变量的变差除以观测值的数量 N,并将标准误定义【Standard Deviation】为方差的平方根。

一变量的变差可以归之于系统力(如影响感兴趣的变量的其他变量)或误差【Error】(如测量该变量的误差)。许多统计技术,如回归【Regression】和方差分析【Analysis of Variance】,试图分解因变量【Dependent Variable】上变差的来源。在回归分析中,总变差被分入来自回归的变差和可归之于误差的变差。例如,如果所有的数据点都落于线性回归【Linear Regression】的直线上。那么,所有的变差都归之于回归,而没有变差来自误差项。回归中 R 方【R-Squared】的值是归之于回归的变差同因变量上的总变差之比。在方差分析中,总变差被分解入归之于各个自变量的部分,可能归之于自变量的交互效应【Interaction】的部分和未解释的变差。应用这些技术时,研究者试图寻找能最小化未解释的误差项变差的自变量——受限于使用最少可能且有效数量的预测变量【Predictor Variable】的简省

(parsimony)要求。

就非数值变量而言,变差涉及检查值的散布性。例如,变差比【Variation Ratio】即是几个可用量度之一,可用以估计类别数据中观测值的散布度。

——Herbert F. Weisberg
(赵锋译校)

参考文献

Blalock, H. M. (1979). *Social statistics* (rev. 2nd ed.). New York: McGraw-Hill.

Lewis-Beck, M. S. (1995). *Data analysis: An introduction.* Thousand Oaks, CA: Sage.

变差系数(Variation [Coefficient of])

一个变量的变差系数量度了一个连续变量的标准差【Standard Deviation】与同该变量样本均值的比。保罗·艾利森(Allison,1978)提出变差系数可用于比较两个或多个组别的相对变差。这一系数通常标记作 *CV*,并定义为

$$CV = \frac{s_x}{\bar{X}}$$

除了传统的样本变差量度,如样本方差和标准误外,研究者还可使用多个组别的变差系数,参照某个变量,比如说 *X*,来比较它们的相对变化程度和同质程度。*CV* 上的较高值表示该组别有相对更大的异质性;相反,*CV* 上的较低值则意味着该组别更加同质。

应用变差系数的基本理念是传统的变差量度,如方差【Variance】(或同等的标准差),是标尺-依赖的(scale-dependent)。从根本上说,如果数据被再标尺,方差和标准差也会被再标尺。例如,如果将某个变量的值倍增,亦即倍增其方差,其标准差也随之倍增。如果将年度收入的标尺从元(如36 000元)变到千元(如 36.0 千元),那么,标准差也将相应缩减 1 000 倍(如55 000~5.5)。

标准差和方差的标尺依赖性使得在用它们作跨组比较时会产生错误。变差系数给出了此问题的一个合理解。因为 *CV* 定义中的分子(即 *X* 的标准差,s_x)和分母(即 *X* 的均值,$\bar{X}$)都对标尺敏感,所以当计算变差系数时,标尺的因素即被排除。由此,*CV* 是标尺不敏感的,并且通常会报告成一个比值或一个简单的十进位数值。

为了说明应用 *CV* 作跨组比较的优点,可以假设有一研究者关注两个社区在年收入上的统计同质性。在一个社区里,收入均值是 26 000 美元,标准差为 3 000 美元。另一个社区中,平均收入是 39 000 美元,标准差为 5 000 美元。我们发现,收入均值高的同时有大的标准差。因此,在试图比较这两个组别时,估计年收入的标准差相对于收入均值的大小是有益的。这样可得出两个变差系数:第一个社区为 3 000/26 000 = 0.115;第二个社区为 5 000/39 000 = 0.128。注意两个 *CV* 间的差比两个标准差间的差小很多。

尽管变差系数给出了一个相对散布的量度,在社会科学文献中还是很少报告该系数。杰斯珀·索伦森(Sørensen,2002)指出了应用变差系数的几项缺点。他主张将均值和标准差合并到一个量中,有时可能模糊这两个因子的真正效应,假如标准差和均值对同一项因变量有独立的效应。他建议研

究者在应用变差系数时,需要谨慎。艾利森(Allison,1978)也对几个可供选择的变差量度间的比较给出了更多信息,如变差系数和Theil指标。

——Cheng-Lung Wang
(赵锋译校)

*也可参见相对变差量度【Relative Variation (Measure of)】。

参考文献

Allison, P. D. (1978). Measures of inequality. *American Sociological Review*, *43*, 865-880.

Sørensen, J. B. (2002). The use and misuse of the coefficient of variation in organizational demography research. *Sociological Methods & Research*, *30*, 475-491.

变差比(Variation Ratio)

变差比是针对定类变量【Nominal Variable】的离中趋势【Dispersion】的一个量度。它反映了众数【Mode】在多大程度上代表了数据。它可以用不在众数类别内的个案的比例来计算。变差比在0~1变化,越大的值表示变量有越大的散布性:

$$变差比 = 1 - (f_{众数}/N)$$

其中,$f_{众数}$是众数类别的频次,而N是个案的总数。

变差比的下界为0,当所有个案都在同一个类别时,取得该值。因此,零值显示变量上不存在散布性。

当众数的频次为1时,变差比的上界是最大的,意味着每个类别只有1个频次,因此,在该变量上,有完全的散布。这时,方差比为1-(1/N),随着个案数量的增加,其值接近1。例如,可能的9位社会安全码的变差比为1-(1/1 000 000 000)= 0.999 999 999,因为每个可能的社会安全码都是唯一的。

如果说一个带有k个类别的变量呈均匀分布【Distribution】,即每个类别上存在相等的频次N/k。这时,变差比等于1-(1/k)。因此,在均匀分布的情况下,变差比的值取决于变量类别的数量。

变差比作为散布性的一个量度,其优点是便于计算。缺点是忽略了数据中的许多信息,因为它没能将个案的完整分布纳入解释。多样性指标、质性变差指标和基于信息理论的熵量都是受所有类别上个案分布影响的散布量度(Weisberg, 1992, pp. 69-74)。

例如,可以思考宗教归属在不同国家内的分布。在美国,大约60%的人是新教徒,所以,在美国,宗教上的变差比就是0.40。与之相比,在波兰,变差比即是0.10,因为,这个国家有大约90%的罗马天主教徒。因此,变差比显示,宗教上的散布性在美国比在波兰更大。然而,这些值并没有反映出其他宗教在这些国家内的散布程度。

——Herbert F. Weisberg
(赵锋译校)

参考文献

Blalock, H. M. (1979). *Social statistics* (rev. 2nd ed.). New York: McGraw-Hill.

Weisberg, H. F. (1992). *Central tendency and variability*. Newbury Park, CA: Sage.

方差极大旋转(Varimax Rotation)

参见旋转【Rotations】。

向量(Vector)

向量是几何空间中有大小和方向的对象。图示上,向量的量通常由线段的长度表示,其方向则由线段终点的箭头表示。文字上,向量可由其上带有箭形符号表示方向的起点和终点来表示。例如,$\overrightarrow{\mathbf{AB}}$表示向量起于 A 终于 B。

在矩阵代数【Matrix Algebra】中,若一个矩阵仅有一列就被称作列向量。与此类似,若一个矩阵仅有一行则被称作行向量。向量的一个常见应用是在回归分析(见普通最小二乘【Ordinary Least Squares】)中,其中,例如它们用以表示在某个 $1\times n$ 的列向量中,n 个因变量的观测值的集合。特征向量和它们对应的特征值【Eigenvector】也可能被用在回归诊断【Regression Diagnostics】中(参见因子分析【Factor Analysis】)。

——Timothy M. Hagle

(赵锋译校)

向量自回归(Vector Autoregression,VAR)

参见 Granger 因果关系【Granger Causality】。

Venn 图(Venn Diagram)

在一个 Venn 图中,几何图形重叠的区域具有意义,可以促进对相互关联的概念的理解。鲁斯基(Ruskey,2001)纵览了 Venn 图的历史和它们作为一个技术概念的技术细节。肯尼迪(Kennedy,2002)详细阐明了在它们同社会科学研究方法相关的背景下,Venn 图的应用如何加强了对多元回归分析【Multiple Regression Analysis】的理解。

图 1 给出了一个经典的 Venn 图,其中记作 y 的圆表示因变量 y 上的"变差",而记作 x 和 w 的圆分别表示解释变量 x 和 w 上的"变差"。当做 y 对 x 的回归时,图中蓝区加红区表示可由普通最小二乘【Ordinary Least Squares,OLS】公式用以产生 x 的斜率系数估计值的"信息"。如果这一信息表示 y 上的变差由 x 上的变差唯一地解释,所得的估计值就是无偏的【Unbiased】。更多的信息,由更大的重叠区域代表,将产生有更小方差的估计值(无论这一信息的"真实性"!)。在多元回归背景下,若 y 同时回归于 x 和 w,OLS 估计量用蓝区中的信息去估计 x 的斜率系数,同时,用绿区中的信息估计 w 的斜率系数,而舍弃了红区中的信息。因为红区中的信息并不对应于 y 上唯一可

归之于 x 或 w 的变差，所以没有使用它。估计的结果是 x 和 w 斜率的无偏估计值。

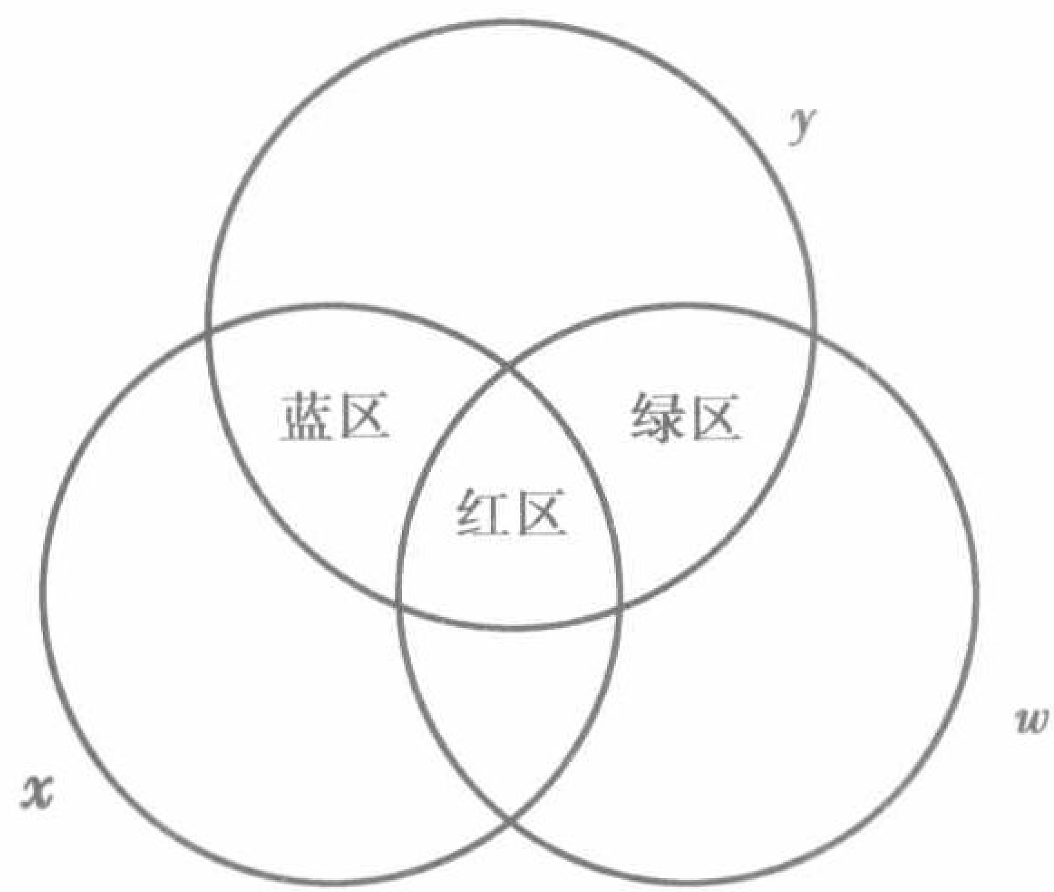


图 1 Venn 图，标注颜色的区域表示可由 OLS 回归生成系数估值的信息

在回归背景下，Venn 图的解释可用于具体说明几项有趣的结果，下面有两个例子：

1.x 和 w 间重叠区域的增加即增加多重共线性【Multicollinearity】的程度。这导致蓝区和绿区的缩减，而红区则相应增大。因为 OLS 公式只应用蓝区和绿区的信息去估计 x 和 w 的斜率，这些估计值仍旧是无偏的，但由于如今只有更少的可用信息，它们的方差则会增加。

2.遗漏相关的解释变量 w 将导致 OLS 公式使用蓝区和红区去估计 x 的斜率。这就会产生由红区的加入而带来的偏倚，但是由于有更多的信息可用，估计值的方差会减少。一个显然的例外是：如果 x 和 w 是正交的（即没有重叠，从而不存在红区），正如一项设计的实验【Experiment】中的情形，即无偏倚生成。

——Peter E. Kennedy

（赵锋译校）

参考文献

Kennedy, P. E. (2002). More on Venn diagrams for regression. *Journal of Statistics Education*, *10*(1) [online]. Retrieved from http://www.amstat.org/publications/jse/v10n1/ kennedy.html.

Ruskey, F. (2001). A survey of Venn diagrams. *The Electronic Journal of Combinatorics*, Dynamic Survey #5 [online]. Retrieved from http://www.combinatorics.org/Surveys/ds5/ VennEJC.html.

口头报告分析(Verbal Protocol Analysis)

口头报告分析是典型的预研究【Pilot Study】，其目的在于，通过对研究主体（特别是消费者）所做决策过程的口头表述进行系统编码，来实施内容分析【Content Analysis】(Kuusela, Spence, & Kanto, 1998)。

口头报告分析将分析的重点直接放在认知事件（发生在信息刺激的引入和决定结果之间）的序列上。通过设置实验操作或者仅设置一项任务，引入选择情境，再引导受试对象口头报告所有的涉及了任务判断/选择的个人想法，由此收集到数据。这些想法通常会被录音，接着被转录，再被划入一个任务-相关的陈述序列（报告片段）中，最后使用一个带类别的编码【Coding】方案，如对选项的配对比较，予以内容分析。因此，反映了短期记忆过程和内心深思熟虑的想法会有助于判断以下问题，如选择的标准、决定的策略，以及规律性。呈现基本信息过程的不同侧面的报告片段，也可定性和定量地予以分析。

与其他信息处理方法相比，口头报告的每单位资料可以提供更多的信息（Arch, Bettman, & Kakkar, 1978）。而且，通过追踪决策过程，口头报告还可作出解释。在没有很好的理论来指导对目标过程的调查时，这一方法在探索性研究中很有用。此外，报告

片段的简单编码可用以计算与理论问题有关的描述性统计量，例如，人们在何时以及多频繁地使用了非补偿性决策规则。

这一方法有些弱点。提供口头报告的需要可能改变刺激中被关注的信息内容（特别是，在可视呈现条件下，可能引起注意力缺乏），以及/或者利用被试对象可得的认知资源（混合过去和当下，产生精度和效度问题），从而根本地改变用于作出决策或判断的潜在过程。同样，这里还存在着传达误差、委托（commission）误差和省略的错误。由于某个不准确的词汇或者在有效回应能力上的不充分训练，就会出现传达不准确。当受试主体为了让研究者留下印象认为他们的行为看起来更合理，而系统地误报他们的认知过程时，就会发生委托误差。当希望口头报告是全面的，但受试主体可能歪曲或省略了特别的要素（诸如眼睛的固着模式等提示性线索可能有帮助）时，遗漏即出现。当报告分析被用于确定人类行为是否受自我-生成的规则管理时，此类问题可以得到部分的缓解（Hayes，White，& Bissett，1998）。为把握潜在的认知过程，人们可以衡量下述三个标准：

1.相关性。参与者应当谈论手头的任务，而不涉及无关的议题。

2.一致性。口头表达，要切题，应当同它们之前的口头表达有逻辑上的内在联系。

3.记忆。任务执行期间留意到的信息的子集应当被记住。

报告引导方法在以下方面有区别：

- 情境或决策任务同何种报告资料被收集有关；
- 方向和引导的总量由访谈员规定；
- 报告收集的时机（并行的与回顾的）。

比回顾报告产生更多报告片段的并行报告能反映出信息处理的方式（何种、何故、怎样）。因为焦点在于过程，所以这个方法适合于预研究。在理想情况下，回顾报告是在任务完成之后立即收集的，这时大部分的信息仍然在短时记忆中，可直接回想起来。研究显示，这时受试主体通常会关注最终的选择，因此，这一方法适用于那些关注任务结果的研究。

——Pallab Paul

Kausiki Mukhopadhyay

（王玥译　赵锋校）

参考文献

Arch, D. C., Bettman, J. R., & Kakkar, P. (1978). Subjects' information processing in information display board studies. In K. Hunt (Ed.), *Advances in consumer research, Vol. 5* (pp. 555-559). Ann Arbor, MI: Association for Consumer Research.

Hayes, S. C., White, D., & Bissett, R. T. (1998). Protocol analysis and the "silent dog" method of analyzing the impact of self-generated rules. *Analysis of Verbal Behavior*, *15*, 57-63.

Kuusela, H., Spence, M. T., & Kanto, A. J. (1998). Expertise effects on prechoice decision processes and final outcomes: A protocol analysis. *European Journal of Marketing*, *32*, 5-6, 559.

确证（Verification）

确证是指使用经验数据、观察、检验或实验来确认某一假设的真实性或具有理性正当性。科学信念必须由经验数据予以评估和支持。这要求什么呢？在讨论科学方法时，两个概念是基本的：真实和证明正当（理据）。如果一个假设同世界所示的方式

相符合,它就是真的。证明正当必然涉及我们根据的基础,而正是以此基础,才能相信某一给定陈述为真。如果某一假设得到了一组经验和推论的支持,那么它是合理地被证实的。理论上,一个陈述经合理地证实的事实应当使这个陈述为真具有可能性。[这种处理明显地呈现了贝叶斯理论对科学推论的适用。问题不得不涉及合理可靠性(rational credibility)从一组信念到另一组信念的转移。]哲学家已引入了一些概念和理论,并以此为基础,去分析经验数据和科学假设之间的逻辑和推论关系。这些概念和理论包括观察、证伪、确认,以及实验【Experiment】方法。这些延伸的、复杂的争论的最重要的后果之一是下述推定,即理论【Theory】不可能是"证明为真的",但是它们可能是"经确认的""经证实的"或"证明为假的"。

科学推论的一般问题可用下面的术语予以形式的表达:给定一组证据 E 和一个假设【Hypothesis】或理论 T,我们如何衡量给定 E 是 T 的证明正当(warrant)?存在着回答这个问题的逻辑基础吗?(对一个科学方法的理论而言,这没有界定其全部的任务,因为科学方法还应当给我们一些指引,以发现一组相关的证据。)对一个科学假设而言,依此方法,它极少能经得起直接的和确定的经验确认:给定 E,H 确定是真的。此即,极少情况下,存在一组有限的观察,其足以确立某个给定科学假设的真实性。所以如此的原因有二:其一,因为科学假设正常情况下涉及直接观察不到的实体、机制或过程;其二,因为假设和理论正常情况下要作出普遍断言(规律),而这些断言超出任一组数量有限的观察。反而,在正常情况下采用间接归纳或假设-演绎的形式去支持假设:给定 E,H 可能为真。

假定 H 推论隐含着 O,同时,一组证据 E 包含"非 O"。此种情况下,E 证明 H 为假。在这些假定下,证据组包含了表明 H 为假的观察。(例如,设定假设 H 必得出,当被加热到 500 ℃时,这种金属即熔化。我们做相关试验,金属没熔化,从而断定 H 是假的。)假定现在 H 包含了一系列观察 O_1, $O_2, \cdots, O_n$,而 E 包含 $\{O_1\}$(此即,发现理论的观察结果中的一些为真)。E 确认 H 吗?在多大程度上?这是归纳逻辑和科学方法的基本问题。对一个理论的显著性检验和确认由何组成?几项逻辑点是重要的。没有有限数量的观察清单能穷尽地确认一个含有普遍概括的理论,而且,不同类型的附加证据对理论的可靠性有非常不同的增量效应。例如,同一类型的附加观察比一个附加的、意外观察有更小的认识论权重。

何谓一项观察?这是科学哲学的一个主要的争议源。在科学哲学中,很久以来就承认,观察和理论之间并没有清晰且永久的区别。实际上,所有的科学观察都是理论-承载的。但是基本观点是,基于可信的数据收集技术(对其我们可以赋予高度合理的可靠性),一项经验观察即一条科学信心,即相信感官证据和陈述的真值条件间具有一种相对直接的关系。此处,比如,让我们考虑一下直接的感官观察、使用工具的观察、访谈、价格数据的记录等。其中大部分既不是"直接的"也不是"确定的"。但是,它们也相对地免予受当前有争议的理论的困扰。这一描述提出了一项经验确认(empirical confirmation)的方法,而该方法可以被描述成"系鞋带"法,其中某些信念的可靠性通过赋予其他信念暂时的可靠性而得到增强;然后,根据更广泛的理论信念集,我们转回来再估计暂时信念的可靠性。

在过去的世纪里,针对科学假设的经验评估,已给出了几种一般的方法。第一种主要是经卡尔·亨普尔深入探研的想法,此即,我们应当引出某一理论的推导结果,使用观察和工具来评估其中一些后果的真实

性，再根据观察后果的量，赋予理论一定程度的证明正当。这一方法构成了假设-推演的确认理论，还阐明了实验方法的逻辑基础。第二种是由卡尔·波普尔提出的想法，此即，科学方法主要通过勤奋且严肃的努力致力于拒绝科学假设。研究者需要指出最不可能预见的假设的后果，并全力以赴去证实这些预见是没有证据支持的——由此问题中的理论“证明为假”。仅当一个理论或假想已经依此方法成功经历了各种严肃检验后，才具有相信它的合理基础。两种方法在逻辑上是相似的，即它们都试图根据假设的推导结果的真或假，赋予某一假设一定程度的经验的证明正当。但是，关于如何实施确认和检验的潜在假设却相当不同。确认理论在很大程度上假定，经由真实结果的逐渐累积，证明正当的程度会增加；而证伪理论假定，只有一个我们没能拒绝问题中的假设的结果出现时，证明正当的程度才增加。两个方法都带来了大量的批评讨论。关于经典的确认理论的批判反思指出，提出一组能够用以增强理论的证明正当程度的，某一理论的推导结果的纯形式的详细说明，不仅是困难的甚至是不可能的。关于证伪理论的进一步讨论也指出，在发展科学理论的过程中，太容易发现反常现象，同时，相关讨论还试图根据“科学研究纲领”理论，给出对科学信念的一个更具历史性的适当说明。

——Daniel Little

（王玥译　赵锋校）

参考文献

Brown, H. I. (1987). *Observation and objectivity*. New York: Oxford University Press.

Glymour, C. N. (1980). *Theory and evidence*. Princeton, NJ: Princeton University Press.

Hempel, C. (Ed.). (1965a). *Aspects of scientific explanation, and other essays in the philosophy of science*. New York: Free Press.

Hempel, C. (1965b). Confirmation, induction, and rational belief. In C. Hempel (Ed.), *Aspects of scientific explanation*. New York: Free Press.

Lakatos, I. (1978). *The methodology of scientific research programmes: Philosophical papers. Vol. 1*. Cambridge, UK: Cambridge University Press.

Laudan, L. (1977). *Progress and its problems: Toward a theory of scientific growth*. Berkeley: University of California Press.

Popper, K. (1965). *Conjectures and refutations: The growth of scientific knowledge* (2nd ed.). New York: Basic Books.

理解（Verstehen）

这个德语词常译作“解释”，且常与“说明（erklaren）”一词相对。这一组对立的词用于19世纪德国对人文科学【Geisteswissenschaften】的特征之讨论，人文科学同自然科学相对立。在此背景下，理解常指从行为者的内心去理解，而不是通过外在行为的研究，同时，它也常被视作人文科学的独特方法。同样重要的是，它产出的是一种不同于自然科学产生的知识。

这一想法的源头可能是维科的提议，即我们能够对其他人有更深层次的理解，包括那些活在过去的人，而我们不可能对物理现象有此种理解。这是因为我们具有共同的人性。然而，此种人性不是一组固定的特质，而是一种想象的能力，一种既能产生也能被用于理解过去的思维模式和生活方式的能力。维科的著作在18世纪中被大大忽略了，但却被德国的浪漫主义拾起。

狄尔泰是19世纪最重要的关于理解的理论家。在漫长的学术生涯中,他消耗了其中绝大部分努力去澄清和发展这一概念,使其成为社会和历史科学的关键方法。他最早的系统表述将理解描述成一种心理学过程,即历史学家想象地把他们自己投入到过去人们的生活中重新创造出他们的经验,并以同样的方式去把握被具体地表达在文本和人造物中的意义。他相信,需要一门哲学心理学(philosophical psychology)或诠释学【Hermeneutics】,因为它将为编史(historiography)提供基础,而他则开始创立它。狄尔泰后来对理解的说明将其描述成文化解释,而不是一种主观过程。它涉及根据构成所要理解的事物背景的客观文化意义,确定何种内容要被理解。同这里相关的狄尔泰的理论部分是承认历史多样性而抵制方法论的实在论【Relativism】。

犹尔泰的想法在社会科学领域中已具有相当大的影响,特别是通过韦伯的理解社会学(参见理想类型【Ideal Type】)。不过,他们也遭遇到严峻的挑战。在20世纪,实证主义者【Positivism】论说,充其量理解也只能生成仍需检验的假设,因为它不包括任何严格形式的验证。另一方面,也有来自伽达默尔的批判,而他提出一项被他称为哲学诠释学(philosophical hermeneutics)的新想法。如果说狄尔泰视理解为一种对社会研究的科学特征具本质性的方法,伽达默尔则将它换为人在世间(human-being-in-the-world)的一个基本方面。从这一立场出发,它没有就理解(understanding)的基础形式提出不同意见,同时,它也没有一种包含程序性的、固定特征的方法。随后,结构主义【Structuralism】和后结构主义【Poststructuralism】中的某些版本则从各自的角度质疑理解的可能性。

——Martyn Hammersley

(王玥译　赵锋校)

参考文献

O' Hear, A. (Ed.). (1996). *Verstehen and humane understanding*. Supplement to *Philosophy*, Royal Institute of Philosophy Supplement 41. Cambridge, UK: Cambridge University Press.

Palmer, R. E. (1969). *Hermeneutics*. Evanston, IL: Northwestern University Press.

Truzzi, M. (1974). *Verstehen: Subjective understanding in the social sciences*. Reading, MA: Addison-Wesley.

素描技术(Vignette Technique)

素描由呈现给研究参与者的刺激构成。它们的目的在于选择性地描绘出现实的不同方面,从而要求参与者对此予以回应。它们采用多种形式,包括书面的和口头的叙事、可视图像、影像和声音。

像任何研究方法一样,素描只能解决清楚界定的研究问题,同时,它们的形式和应用也受所提的研究问题、研究的主题及参与群体的类型的引导。场景的简单性可以帮助辨识、澄清及理顺真实世界过程的复杂性。它们也给予研究参与者标准化的背景,以理解人们对特定情境的反应。

思考一下休斯(Hughes,1998)探索毒品注射者的感染风险行为的定性研究【Qualitative Research】,它给出了应用素描方法的一个例子。先前的研究表明,毒品注射者的风险行为受到所处背景的极大影响。素描被用来部分地把握其中的某些情境。

研究者创作了一个简短的故事书形式的素描,叙述的情节包括两名假设的毒品注射者面临的风险行为。研究者将毒品注射者对素描的回应设置在一个深度访谈提纲【Interview Guide】中,从而展示了素描在多方法研究中的效用。休斯观察到,毒品注射者与素描有很好的关联,其中有些人喜欢这个故事,并有自信心去预计素描人物的可能后果。此外,参与者还从中得出一系列视角,为素描人物、他们的伙伴和他们自己作答。这些不同的视角有助于最小化研究的潜在敏感话题【Sensitive Topics】和社会期望偏倚【Social Desirability Bias】。在应用时,素描还暴露了某些包含在素描情节中的风险行为的构成要素,其中包括与人们所相信的毒品注射者生活中可能有的现实反应相比,人们认为应当发生什么事情。

素描不可能完全把握人们生活的现实,而且,像所有研究方法一样,它们的应用取决于研究目的。对素描的方法论争论集中在下述差别,即人们在真实生活情境中实际可能做的,与勾勒于"素描世界"中的真实生活的选择呈现诱导出的参与者反应之间的差别。

有关该方法的批评,包括帕金森和曼斯蒂德(Parkinson & Manstead,1993)的论说,素描资料只能在具体背景(人们对特定情节的反应)中理解,而不容许抽象概括,以理解真实生活。其他评论者,如洛曼和拉津(Loman & Larkin,1976),主张某些素描形式,如影像和电影,相对其他形式来说,如书面叙事,更贴近于真实生活。尽管存在诸多批评,素描仍在增加对社会世界的洞见中扮演着有价值的角色,也很适合多方法研究。在整个社会科学领域中,它们都有助于理解人们的感知、信念、态度和行为。

——Rhidian Hughes

(王玥译　赵锋校)

参考文献

Hughes, R. (1998). Considering the vignette technique and its application to a study of drug injecting and HIV risk and safer behaviour. *Sociology of Health and Illness*, *20*, 381-400.

Loman, L. A., & Larkin, W. E. (1976). Rejection of the mentally ill: An experiment in labeling. *Sociological Quarterly*, *17*, 555-560.

Parkinson, B., & Manstead, A. S. R. (1993). Making sense of emotion in stories and social life. *Cognition and Emotion*, *7*, 295-323.

虚拟民族志(Virtual Ethnography)

虚拟民族志这个词表示对以信息和沟通技术为中介的、文化活动的参与观察【Participant Observation】。这些网络活动本身被看成以自身特有的方式进行的文化交换(cultural interchanges),因此,民族志学者可以参与其中进行观察,并对其文化形式和过程作整体描述,就如同在面对面的情景中进行的民族志工作。民族志学者可能研究认同的形成、地位等级结构、权力关系、文化价值及行为的准则。民族志可以描述一个网络环境如何被参与者变成一个独特的文化空间,例如科雷尔(Correll,1995)描述的女同性恋吧(电子吧)。虚拟民族志可以使用定性为基础的身临其境的方法或者他还可以扩展到对互动模式更加复杂的图示化和量化。实地调查地点可能是一个不同步的讨论论坛,例如,一个新闻组或公共留言板,或者是一个同步的环境,例如,聊天室或

多用户环境(文本为基础或有图形的)。在每个情况下,研究者可以保留信息或互动的记录,以及自己的观察和反应的田野笔记。

在一些网络环境中,研究者在没有参与者的许可或参与者不知情的情况下,也能够观察互动,这就产生了伦理上的困境。一些研究者可能主张:在投入一个观察研究前,应当仔细考虑所讨论的话题的敏感性,以及应当对参与者所可能使用的词语有合理的预期。运用虚拟民族志的其他问题包括:无法获得信息人线下生活的确切证据;分析仅局限于那些积极的参与者,而那些只阅读信息而不发信息的人(潜水的人)的反应是看不见的(Paccagnella, 1997)。如果目标是描述网络文化本身,这不是一个问题,但是如果目标是要说明更广泛的文化现象,那么这就是有问题的。一种借鉴更加公开的反身性【Reflexivity】的方法强调,民族志学者可以从他们自己通过技术与信息人互动的经验中学到许多,同时,这一方法更加强调理解以技术为中介的在场(the technological mediation of presence)的经验。

虚拟民族志与在线社区视角(the online community perspective)紧密关联,而该视角认为,建于在线基础上的社会形态,例如互联网新闻组,本身可以被看成社区(Baym, 2000)。尽管缺少面对面的互动,以及潜在的长期责任的缺失,线上的社会形态对于参与者是有深层意义的。海因(Hine,2000)认为,在线社区视角鼓励对民族志方法的保守诠释。虚拟民族志的另一个概念为空间创新方法论的方法提供了可能性,即利用民族志的敏感性去探索多种多样的线上和线下环境间的文化联系。这样,虚拟民族志就成了一种有着广泛应用前景的方法论,可以包含更广泛的研究主题,而不只是网络文化。

——Christine M. Hine

(王玥译　赵锋校)

参考文献

Baym, N. (2000). *Tune in, log on: Soaps, fandom and online community*. Thousand Oaks, CA: Sage.

Correll, S. (1995). The ethnography of an electronic bar: The lesbian cafe. *Journal of Contemporary Ethnography*, *24*(3), 270-298.

Hine, C. (2000). *Virtual ethnography*. London: Sage.

Paccagnella, L. (1997). Getting the seats of your pants dirty: Strategies for ethnographic research on virtual communities. *Journal of Computer Mediated Communication*, *3*(1). Retrieved from http://www.ascusc.org/jcmc/vol3/issue1/ paccagnella.html

可视研究(Visual Research)

可视研究涵盖了一系列的方法和实践,而它们都将可视媒体和技术用于研究的所有阶段。自1990年代早期以来,在社会科学研究中,可视研究应用的迅速增多有两个原因:第一,社会科学领域中理论转换使得研究者承认可视材料并不一定比口头或书面资料更主观;第二,质量不断提高、成本不断降低、易用性不断增强的新可视媒体和技术唾手可得。因此,越来越多的研究者在他们的研究工作中使用了视频、摄影、绘图和多媒体。

尽管可视研究这一词已变得通俗,但是可视研究很少是纯视觉的,而是在做研究和将研究成果呈现给他人的过程中,结合了视觉的、口头的和书写的方法。因此,可视方法不应脱离既有的方法孤立使用,而是一组用以认识和说明日常生活中和人类知觉过程中视觉经验、知识和交流的重要性的工具。可视研究包括以下内容:

1.分析现有可视图像的内容、生产过程

或使用方式,包括定性或定量的内容分析【Content Analysis】(Rose, 2001)。

2.将可视图像的生产作为研究项目的一部分,以可视的方式记录活动和事件,或者与被调查者合作来生产图像,以呈现特定的观念、理解或世界观(Banks, 2001; Pink, 2001)。

3.在访谈中使用图像,从而引起被调查者的反应(Banks, 2001; Pink, 2001)。

4.对时间和活动的可视观察(Emmison & Smith, 2000)。

通常,个别研究者会在一个研究项目中混用可视方法与其他方法,如结合书面文本的分析,深度访谈【In-Depth Interview】或参与观察【Participant Observation】。

根据所用的媒介和研究目标,可视研究有多种呈现方式。这包括在图书出版物中使用照片;摄影展示或摄影集;民族志研究中的影像【Film and Video】;还有发布在网络、CD-ROM 或 DVD 上,使用多媒体交互式的超媒体项目。

可视研究的历史

在 20 世纪初期,可视方法首次用于社会学和人类学。这主要包括早期关于欧洲贫民、海外殖民地的"民族志摄影",以及在人类学和探险考察中制作的关于其他文化的电影。这一阶段内许多科学图像的生产近来已受到严厉的批评,并被视为他者化过程(exoticising)及压迫性殖民工程的一部分。然而,这些图像毕竟是有价值的研究记录。20 世纪末,两个里程碑式的项目标志着可视研究的进一步发展。《北方的纳努克》(*Nanook of the North*, 1921)被称作第一部民族志电影。为了表现北美爱斯基摩人日常生活的挑战和欢欣,纳努克设置了许多有争议的场景,而这些争议在整个 20 世纪的民族志电影制作中一直延续。1942 年,米德和贝特森的摄影研究《巴厘人物》(*Balinese Character*)发表。为了试图以可视的方式记录巴厘岛人的行为,这个作品承载着那个时代社会科学家们的目标——生产出人类活动的、客观的可视资料。而可视图像是否是客观的问题成了 20 世纪争论的关键议题。

1980 年代和 1990 年代,不断变化的理论氛围借势促进可视研究的发展。关于客观性【Objectivity】是否真的能够实现,这一观念在社会科学领域,特别是在定性研究者中,遭遇挑战。反而,反身性【Relexivity】的观念——研究者对他们的主体性的自我意识以及这点如何影响产生的资料的类型——在定性研究【Qualitative Research】中变得愈益重要。在这种氛围之下,有一点变得很清楚,即可视资料同其他资料比,不再是太过主观而不能产生可靠知识的材料,同时,研究者可以为在他们的研究和作品发表中运用可视方法提出有说服力的论证。

当代可视研究的应用

当代的可视研究者主要使用摄影、视频、绘图和超媒体于他们的作品中。关于哪一种可视方法或媒体是最合适的,通常没有严格的规则。通常,方法的设计是专门的,以适应更广泛的研究目的,同时也要适于研究者与之工作的特定人群。例如,班克斯叙述了在印度的一个节日期间,他如何把摄影作为参与观察的一部分。他的被调查人让他给关键的人和活动拍照,从而向他表明在这一事件中什么事物对他们来说是最重要的。

通过这种方式,他证实他自己对这些人和行动的重要性的理解是正确的(Banks, 2001, pp. 45-47)。在研究西班牙斗牛时,我体验了相似的"知情人-导演的摄影"——我的知情人让我拍照片,由此我认识到我自己以及他们关于斗牛的理解(Pink, 2001)。这种情况下产生的照片也可能用于照片-诱

引的访谈(photo-elicitation interviews),而这种访谈容许研究者和知情人讨论可视材料的意义和内容。深度访谈也可能成为生产图像的背景。例如,在研究人与他们的家的关系时,我用录像记录了整个访谈过程,而访谈由对每个知情人的家巡回构成,同时,摄像以我的问题清单为指导,由被调查人主导。这使得我可以与我的知情人共同探索他们的家的可视方面,以及不同的对象、图像和他们在家中占据的不同区域的意义。在这些例子中,可视方法不是作为同既有方法不同的特殊方法而使用;相反,它们被结合到参与观察或深度访谈中。

然而,这种结合完全不是简单的。应用可视方法产生了新的伦理问题。不像仅以书面形式呈现的研究,知情人的匿名性在可视材料中无法得到保证。因此,这些研究要求在研究开始之前,与研究中的被调查人和其他利益相关者充分商讨,仔细商议。

可视研究方法在研究者中逐渐变得流行。它们的应用和理解必定在社会科学的不同领域中稍有不同,因为,在每一领域中,它们都充满了相关学科的理论和方法论的特性。

——Sarah Pink

(王玥译　赵锋校)

参考文献

Banks, M. (2001). *Visual methods in social research*. London: Sage.

Emmison, M., & Smith, P. (2000). *Researching the visual*. London: Sage.

Pink, S. (2001). *Doing visual ethnography: Images, media and representation in research*. London: Sage.

Rose, G. (2001). *Visual methodologies*. London: Sage.

自愿受试者(Volunteer Subjects)

在1960年代和1970年代早期,由于发现自愿者的回答可能不同于一般总体中非自愿者的回答,自愿受试者(自愿参加调查研究的人)成为关注的焦点。拉尔夫·罗斯诺和罗伯特·罗森塔尔(Rosnow & Rosenthal,1970)提出,源自自愿者偏倚的系统误差【Systematic Error】,尽管不易识别,仍可能对实验或非实验的社会科学研究产生微妙影响。通过实验研究,以及对既有文献的系统梳理,罗森塔尔和罗斯诺(Rosenthal & Rosnow,1975)证明自愿者和非自愿者在至少17个特征上存在差异(包括教育、社会阶级、智力、社交性及社会认可动机等),自愿受试者对研究需求特征【Demand Characteristics】既过度敏感又过度适应。

据罗斯诺和罗森塔尔(Rosnow & Rosenthal,1997,pp.89-91)的论证,这些发现对社会科学家至少有3项实践上的含义。第一个含义只使用自愿受试者可能会削弱研究中的科学模型的解释力,因为自愿者效应通常不是模型中一个被清晰表述的变量。第二个含义与将研究结果概论【Generalization】到总体有关,因为总体一般包括非自愿者。第三个含义正如贝尔蒙报告所陈述的属于伦理的考量。贝尔蒙报告是一项为保护生物医学和行为科学研究中的研究参与者权利和福利的指导方针,于1974年由美国的一个国家委员会(National Commission, 1979)负责制定。该指导方针指出,在大多数研究性的调查中,只可使用自愿受试者。实际上,此方针

的提出意味着在研究设计【Research Design】中,不可能消除自愿者问题,只能去控制或修正它。即使如此,精确估计偏倚的程度是一个困难的任务,因为难以知晓总体参数。大多数情况下,科学家能做的最多的是去推测偏倚的方向,并据此解释结果。值得庆幸的是,罗斯诺和罗森塔尔(Rosnow & Rosenthal,1997, pp.95-98)列举了我们对自愿者的信度特征及自愿参与各种研究信度的知识,而这些知识可用于偏倚的推测。

与自愿受试者相关的一个问题是伪自愿者,即一个口头上表示参加,但实际没有赴约的自愿者。早期的研究已表明,研究结论包括那些自愿者和非自愿者之间差异的研究,很显著地依赖于伪自愿者是否被包括进自愿组或非自愿组。通过对研究报告的元分析,罗森塔尔和罗斯诺(Rosenthal & Rosnow,1975)发现,平均而言,接近 1/3 的自愿者没有遵守约定;更多的近期研究确认伪自愿者比率的中位数【Median】从 1975 年以来没有显著变化。

——Ram N. Aditya
Ralph L. Rosnow
(王玥译 赵锋校)

参考文献

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research.* Washington, DC: Government Printing Office.

Rosenthal, R., & Rosnow, R. L. (1975). *The volunteer subject.* New York: Wiley.

Rosnow, R. L., & Rosenthal, R. (1970). Volunteer effects in behavioral research. In K. H. Craik, B. Kleinmuntz, R. L. Rosnow, R. Rosenthal, J. A. Cheyne, & R. H. Walters (Eds.), *New directions in psychology* (pp. 213-277). New York: Holt, Rinehart and Winston.

Rosnow, R. L., & Rosenthal, R. (1997). *People studying people: Artifacts and ethics in behavioral research.* New York: W. H. Freeman.

W

Wald-Wolfowitz 检验(Wald-Wolfowitz Test)

Wald-Wolfowitz(1940)检验是一种快速、易行的用以检验两个相互独立的随机样本抽取自同一总体【Population】的虚无假设【Null Hypothesis】的方法。该检验完全是普遍适用的,并不专门针对总体间差异的特定类别。

为实施检验,两个样本数据集,比如说,m 个 X 样本和 n 个 Y 样本被合并在一起,并按照从最小值到最大值排列成由 $m+n=N$ 个观测值组成的单一数列,同时,标识出哪个样本生成了该观测值。然后,用 X 和 Y 来替代数列上的观测值,以表示各观测值的总体来源。如果样本 X 和样本 Y 在该合并数列上充分混合,就没有理由认为两个总体有任何形式的差别。如果样本 X 和样本 Y 显示出一个模式,这就表示存在着一个差别。例如,如果样本 X 中的多数或全部先于样本 Y 中的多数,这就表明,Y 总体比 X 总体倾向于有更多量大的值。

在这个由 m 个 X 样本和 n 个 Y 样本合并的数列中,一个样本 X 的链被定义为一个 X 样本的序列,而在这一序列之前或之后,有一个或多个 Y 样本,或者根本没有样本。Y 样本的链有相似的定义。Wald-Wolfowitz 检验统计量 t 是样本 X 和样本 Y 的链的总数。当样本 X 和样本 Y 混合充分时,会有大量的链,因此,t 的大值趋于支持同一总体的虚无假设。少量的量表示存在着模式,也就是总体不相同。因此,对于小的 t 值,应当拒绝虚无假设。有关 t 在虚无假设为真的情况下的精确分布表及/或 t 的临界值【Critical Value】可以在某些非参数统计学书籍中找到。就较大样本量(m 和 n 不少于10)而言,基于标准正态分布【Normal Distribution】,可以有一个近似的检验。下面的统计量是 t 的函数,还加入了一个连续的修正项,以增加精确度:

$$Z=\frac{T+0.5-1-\dfrac{2mn}{N}}{\sqrt{\dfrac{2mn(2mn-N)}{N^2(N-1)}}}$$

例如,假设有一个由 6 名政府经济学家(G)和 7 名学院经济学家(A)组成的随机样本,他们都被要求预测同上年相比 CPI 总的百分比变化。数据如下:

G:3.1,4.8,2.3,5.6,0.1,2.9
A:4.4,5.8,3.9,8.7,6.3,10.5,10.8

两组预测值来自同一总体吗?

我们按下面的方式将数据合并成一个单一数列,其中政府数据值以斜体表示。

0.1,*2.3*,*2.9*,*3.1*,3.9,4.4,
4.8,*5.6*,5.8,6.3,8.7,10.5,10.8

以G代表政府，A代表学院，这一排列可以写作GGGGAAGGAAAAA，共有4个链，即$T=4$。在吉布森（Gibbons，1997，p. 483）表中，当$m=6$，$n=7$时，左尾p值【p Value】为0.043。使用大样本近似，我们可得$Z=-1.73$。从标准正态分布表可查知，其左尾p值为0.041 8。这个值非常接近于精确值。在0.05的置信水平上，我们拒绝虚无假设，同时推定两个总体分布不相同。

注意并不必须以实际的测量值来实施该检验，只要能够形成合并的数列即可。Wald-Wolfowiz两样本检验已经包含在统计软件STATXACT中。

——Jean D. Gibbons

（赵锋译校）

参考文献

Gibbons, J. D. (1997). *Nonparametric methods for quantitative analysis* (3rd ed.). Columbus, OH: American Sciences Press.

Wald, A., & Wolfowitz, J. (1940). On a test whether two samples are from the same population. *Annals of Mathematical Statistics*, *11*, 147-162.

加权最小二乘（Weighted Least Squares）

加权最小二乘（WLS）是普通最小乘【Ordinary Least Squares，OLS】回归的一个扩展，即根据某些准则，比如估计值精度或误差项方差上的差异，为每个观测值加权。在常规的OLS中，斜率【Slope】的估计值由最小化平方和得到，即$\sum E_i^2$，其中E表示距回归的残差。此即，最小化实测值的平方同回归方程预测值的差值的平方。WLS通过最小化加权的最小乘方$\sum \omega_i^2 E_i^2$扩展了这一模型，其中ω表示应用于每个观测值的权。根据WLS回归得到的估计值同OLS法得到的估计值的解释方式完全相同。在统计推论时，WLS同样有效且同OLS有相同的标准程序（例如，置信区间【Confidence Interval】和假设检验）。

在社会科学中，WLS最常用于当回归分析中出现异方差性【Heteroskedasticity】（非恒定的误差项方差）。与异方差性相关的两个问题，即无效的估计值和有偏的标准误，可以因加权最小二乘的使用得到控制。无论函数的关系形式是线性的还是非线性的，都可以使用WLS。

然而，在执行WLS矫正异方差性之前，人们必须知道误差项方差是系统地同自变量之一相关。因为这点在拟合回归模型之前是未知的，所以必须先行拟合一个普通最小二乘回归，再考察模型的残差，从而得到WLS回归中的权。异方差性一个常见的模式是随着某个自变量X的增加，误差项方差也随之增加。在这种情况下，误差项的方差同X成比例，由此，$1/X$常是用以计算加权最小二乘估计值的一个有效的权。学生化残差的自然对数（the log of the studentized residuals）同拟合值的图示可以帮助人们查明此现象。

应用WLS来矫正异方差性的缺点是需要关于误差项和某个特定自变量间关系的知识。由于人们不是总能查明这点，因此，带稳健标准误的OLS常用作备选方法。不同于WLS，稳健标准误的使用不会改变OLS回归系数，只是增加标准误的大小。换言之，OLS和稳健标准误的使用会增进统计检验，但它不会必然地给出可靠的系数的估计值。其结果是，如果异方差性的模式可以被查明，那么WLS是较带稳健标准误的OLS

更好的选择。

当误差项方差的非随机模式出现在一个时间序列【Time Series】中时，一个相关的方法，广义最小乘方【Generalized Least Squares】就被用来代替加权最小二乘。迭代的加权最小二乘（也称作迭代的再加权最小二乘）也可用于拟合广义线性模型【Generalized Linear Models】。最后，加权最小二乘也用在局部回归【Local Regression】和稳健【Robust】回归中，前者用以得出局部的估计值，后者则用以减少有影响力的个案对回归估计值的影响力。

——Robert Andersen

（赵锋译校）

参考文献

Fox, J. (1997). *Applied regression analysis, linear models, and related methods*. Thousand Oaks, CA: Sage.

Neter, J., Wasseman, W., & Kutner, M. H. (1989). *Applied linear regression models* (2nd ed.). Boston: Irwin.

加权（Weighting）

加权是当数据取自一个样本时，用以估计总体的某个特征值的一个方法。加权一个样本主要通过附加一个乘数因子（称为权或个案权）给每个样本观测值来实现。某个总体特征值的估计值是样本观测值之和，其中每个观测值都与它的权成比例地被纳入计算。从这个意义上说，加权是用以实现样本到总体的统计推论的方便办法。权的常用同义词有加码（raising）、通胀（inflation）或膨胀因子（expansion factor）。这些术语都是描述性的，因为权将样本抬高至总体的水平。

概率样本的加权

某些加权方法用以针对几乎所有样本来分析观测值。即使某个样本的非加权的分析也是一个所有个案权等于 1 的加权方案。根据应用的需要，加权程序不同。在概率抽样【Probability Sampling】的情况下，加权观测值会产生无偏的或接近无偏的总体特征值的估计值，而无偏【Unbiased】意味着根据所有可能样本计算的估计值的平均数等于总体的特征值。例如，根据取自总体为 N 的一个样本量为 n 的简单随机样本，计算总体的无偏估计值，每个观测值的权就是 N/n。一般而言，在一个概率样本中，选取调查单位的概率之倒数常被称为基权（the base weight）。

基权可以作出调整，以对那些无回答的调查单位或在抽样中没涵盖到的调查单位作出补偿。例如，在电话调查【Telephone Survey】中，可以调整基权，以便对无回答的家庭户和没有电话的家庭户作出说明。另一个调整基权的理由是执行一项更加成熟的估计技术，比如，在估计比率和回归估计量时，利用附加数据来增进估计值的精度。

观测样本的加权

对选自没有使用概率抽样的观测样本加权的主要目标也包括减少偏倚，但是在此情况下，选择性偏倚【Selection Bias】是主要的考量。例如，在一项选民的观测研究中，权或平衡因子必须得到发展，以使得加权的样本同已登记的选民总体的年龄、性别和人种相匹配。加权的目的是减少估计值中的偏倚，而这种偏倚可能由于样本同总体分布

不相匹配。

标准化作为一项分析工具,常用于分析流行病学的研究。当样本来自不同成分构成的总体时,为了便于比较,也常使用标准化。基什(Kish,1987, pp. 113-115)论述到标准化作为观测研究中的一种加权方法,可以减少选取偏倚。他说明标准化是一项加权技术,它消除了不等子组样本量的效应,以及由这些不等样本量带来的偏倚。例如,设想有一群相等数量的男性和女性病患就某项疾病获得治疗,但是用以估计治疗效果的样本中,女性比男性多。这时,可以将权加之于抽样的病患,以使得加权后的男性数量等于加权后的女性数量。

历史的发展

在高速计算机得以应用之前,用个案权来加权既费时又易出错。其结果是:自加权样本——其中每个被抽取的单位有相同的选取概率和相同的个案权——非常通行。随着计算机的引入,包含来自观测值数据的机器可读卡可被复制,以便适当地代表以不相等的选取概率选出的单位。例如,如果某些观测值以 1/10 的概率被抽取,而另一些则以 1/20 的概率被抽取,就可以复制对应于那些以 1/20 的概率被抽取的卡。

被抽取单位 i 的基权,ω_i 是选取概率的倒数,而且是 1 与 N 间的一个数。相应地,人们自然会认为被抽取的单位 i 代表了总体中的 ω_i 个单位。尽管这样的思考可能对于理解抽样和加权有所帮助,但是当考虑更高级的加权方法时,这就失去了效力。例如,一个回归估计量是基权乘以一个调整因子,同时,对于某些个案,受调整的权可能小于零。负的回归权不意味着被抽取的观测值代表了总体中某个负量的单位。在这种情况下,将权简单地视作一个用以计算总体的某个估计值的工具可能更为恰当。

例示

假设有一学校的简单分层【Stratified】随机样本被选取。根据在册儿童的数量,总体中所有学校的清单被分作 5 层。在层 h ($h=1:2,\cdots,5$) 中,每个学校的选取概率等于 n_h/N_h,其中 N_h 是层 h 中学校的总数,而 n_h 是层 h 中抽中学校的数量。层 h 中的第 i 个学校的基权是 $\omega_{hi}=N_h/n_h$,即选取概率的倒数。如果仅 r_h 个被选中的学校有回答,那么对基权的加单校正是 n_h/r_h,即层内回答率的倒数。对于层 h 中的第 i 个学生,基权同无回答校正之积即无回答校正权 $\omega'_{hi}=N_h/r_h$。

假如总体中,根据对学校的普查,公立学校和私立学校的总数已知(比如说,X_1 是公立学校的数量,X_2 是私立学校的数量),人们就可以计算后分层(poststratified)或比率估计量。后分层估计量的加权同计算标准化的加权方式非常相似。所有回答的公立学校的校正因子是 $X_1/\hat{X}_1$,其中 $\hat{X}_1=\sum\omega'_{hi}\delta_{hi}$,而当层 h 中的第 i 个学校是公立时,有 $\delta_{hi}=1$,否则即为零。私立学校的因子为 $X_2/\hat{X}_2$,其中 $\hat{X}_2=\sum\omega'_{hi}(1-\delta_{hi})$。于是,用以计算后分层估计量的权记作 $\omega''_{hi}=\omega'_{hi}X_k/\hat{X}_k$,其中,就公立学校有 $k=1$,就私立学校有 $k=2$。因此,无论是无回答校正权还是后分层权都可用于得出总体的估计值,所以上述例子说明加权方案不是唯一的。

——J. Michael Brick

(赵锋译校)

参考文献

Babbie, E. (1979). *The practice of social research*. Belmont, CA: Wadsworth.

Kalton, G. (1983). *Introduction to sampling*. Beverly Hills, CA: Sage.

Kish, L. (1987). *Statistical design for research*. New York: Wiley.

Särndal, C. E., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling*. New York: Springer-Verlag.

Welch 检验(Welch Test)

两个独立样本均值相等的 t 检验【t-Test】对于方差齐性(同方差性【Homoskedasticity】)和每个样本分布的正态性假定之违背是相对稳定的(Wilcox,引自 Algina, Oshina, & Lin, 1994; Winer, 1962)。然而,如果两个样本的样本量和两个总体的方差不相等时,t 检验的结果可能导致错误的判断。如果较小的样本有更大的方差,普通的 t 检验会倾向于接受 H_1 假设;如果较大样本有更大的方差,H_0 则更可能被接受。Welch 检验(Welch,1947)校正了普通 t 检验的相关偏差。如果样本量相等,普通 t 检验和修正后的 Welch t 检验的检验统计量有相同的值,只是自由度【Degrees of Freedom】有区别。一般而言,Welch 的近似值会有较低的自由度。

Welch 检验的检验统计量记作

$$t_{\text{Welch}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

自由度的计算公式为

$$df_{\text{Welch}} = \frac{1}{[c_1^2/(n_1 - 1)] + [c_2^2/(n_2 - 1)]}$$

其中,$c_i = (s_i^2/n_i)/[(s_1^2/n_1)+(s_2^2/n_2)]$,$i=1$ 或 2。

如果两总体方差不相等(异方差性),则应使用 Welch 检验。方差齐性的假设可通过 Levene 的 F 检验【F Test】予以验证。

例示:比较两个组(见表 1)。组 1 由 10 个个案组成,在变量 X_1 的均值为 10.0。组 2 由 40 个个案组成,变量 X_1 的均值为 12.0。

两组的方差显著不同(表 2)。Levene 的 F 检验的值为 8.598 3(100 次模拟的平均数)。

表 1 组的统计量

	组	N	均值	标准差
X_1	1	10	10.026 3	0.943 8
	2	40	11.980 5	2.963 5

注:这是一项模拟研究的结果,其中运行了 100 次的模拟。假定第一组来自均值为 10,方差为 1 的正态分布;而第二组来自均值为 12,方差为 3 的正态分布。

表 2 独立样本检验

		方差相等的 Levene 检验		均值相等的 t 检验		
		F	sig.	t	df	sig.(2-tailed)
X_1	假定等方差	8.598 3	0.012 4	−2.045 4	48	0.046 3
	不假定等方差			−3.517 6	44.981 3	0.001 0

注:这是一项模拟研究的结果。研究设定已经描述于表 1。报告的值是平均数。F 值是 100 次模拟的平均数,F 值的显著水平也是 100 次模拟的平均数(并非真的是值 8.598 3),其他类此。

计算的误差水平为 0.012 4(100 次模拟的平均数)低于所选显著水平 $p<0.05$ 的阈值。因此,应当使用 Welch 检验。Welch 检验的结果记录于表 2 的第 2 行。检验统计量为−3.517 6,自由度为 44.981 3。如果以 $p<0.01$ 做显著性水平,检验统计量是显著的。相反,普通 t 检验的结果不显著。之所

以有这样的差异是因为当较大的样本有更大的方差时,普通的 t 检验偏向于 H_0。这个例子是基于模拟的,此即,确切的检验统计量是已知的。它的值为-3.51,而且对所选 $p<0.05$ 的阈值是显著的。然而,如果两组样本量都增加,组 2 的方差减少,或者个案数被更均匀地分入两个组,两个检验间的差异则消失。例如,如果组 1 由 20 个个案组成,组 2 有 80 个个案,两个检验都会得出有显著差异的结果[普通的 t 检验:$t=-2.9026$ ($p<1\%$);Welch 检验:-4.9746 ($p<0.1\%$)]。

在一般的统计软件,如 SPSS 或 SAS 中,都可执行 Welch 检验。如果潜在的分布变得更长,且样本量减少(比如每组从 20 减少到 10)(Yuen,1974),检验就变得非常保守,即偏向 H_0。同普通的 t 检验一样,如果潜在的分布是(非常)偏态的(Algina et al.,1994),特别由于不同方向上的偏态【Skewness】导致的问题,Welch 检验可能在两个方向(偏向 H_0 或 H_1)上都产生偏倚的结果。取决于偏态的程度和类型,为了得到有效的结果,可能需要有 40 或 80 合计的或总的样本量,在非常极端的情况下,需要 200 个个案。

好的教科书经常建议,如果样本量小,需要检验每个组中观测到的分布是否偏离正态分布(见 Kolmogorov-Smirnov 检验【Kolmogorov-Smirnov Test】);如果正态性假定被违背,则要使用非参数检验(比如,Mann-Whitneyu 或 Wilcoxon 秩和检验)。有时出现在教科书和研究方法书籍中 V 的一个对此建议的错误解释是,总的分布应是正态的。其实,偏离于正态性不必然等同于 t 检验或 Welch 检验的不佳表现。例如,元(Yuen,1974)记录了在潜在分布为均匀分布且小样本(如每组 $n=10$)的情况下,Welch 检验的优越表现。

偏态,而非背离于正态性,似乎是一个重要的影响因素。因此,建议如果合计的样本量少于 200,人们应当作偏态检验;如果数据是偏斜的,则做 t 检验或 Welch 检验,以及非参数检验,并比较两种检验的结果,除非有更多的信息可用。在更小样本的情况下,特异值可能导致异方差性和偏态。因此,应当检查数据是否有特异值。

——Johann Bacher

(赵锋译校)

*也可参见均值差【Difference of Means】。

参考文献

Algina, J., Oshina, T. C., & Lin, W.-Y. (1994). Type Ⅰ error rates for Welch's test and James's second order test under non-normality and inequality of variance when there are two groups. *Journal of Educational and Behavorial Statistics*, *19*, 275-291.

Welch, B. L. (1947). The generalization of student's problem when several different population variances are involved. *Biometrika*, *34*, 28-35.

Winer, B. J. (1962). *Statistical principles in experimental designs*. New York: McGraw-Hill.

Yuen, K. K. (1974). The two-sample trimmed t for unequal populations variances. *Biometrika*, *61*, 165-170.

白噪声(White Noise)

一般而言,白噪声是一个非系统的变量,代表了未测量的属性。这一用语最常用

在其中有自相关【Autocorrelation】(即变量的当期值同前期值相关)的时间序列【Time Series】变量的研究中。最常见的是一阶过程,其中变量在时点 t 的值与在时点 $t-1$ 的值相关。可以以1995—2002年的美国年度失业率为例。不难设想本年度的失业率可以根据前一年的失业率近似地预测。当然,某些时间序列可能显示或经过转换后显示,在任何水平的滞后【Lag】上,都不存在相关。于是,滞后的值就不能用于预测后来的值,变量似乎由被当作白噪声的随机【Stochastic】过程所驱动。如果所研究的变量实际上是来自回归【Regression】方程的残差【Residual】,可能就需要检验它作为白噪声的可能性。巴特利特提出过一种白噪声检验方法。假定回归残差被发现是白噪声,就无须用通常的自相关技术来获取更有效的估计值。

——Michael S. Lewis-Beck

(赵锋译校)

Wilcoxon 检验(Wilcoxon Test)

Wilcoxon(1945)检验,也被称作Wilcoxon 符号-秩检验,是用于单样本情况下均值的学生 t 检验【t-Test】和配对样本情况均值差的配对样本学生 t 检验的非参数对等物。这里只在配对样本的背景下讨论该方法,因为这是最常见的应用背景。

设$(X_1,Y_1),(X_2,Y_2),\cdots,(X_n,Y_n)$是抽取自某个双变量分布的 n 对随机样本【Random Sample】。我们欲推论样本 X 和样本 Y 的总体【Population】间的均值差。第一步是计算每一对中 X 和 Y 的观测值间的差值,$D_i=X_i-Y_i$。因为 X 总体均值与 Y 总体均值间的差值等同于差值总体 $D=X-Y$ 的均值 μ_D,所以有虚无假设【Null Hypothesis】$H_0:\mu_D=0$。

欲实行 Wilcoxon 检验,首先要按从最小值到最大值的顺序排列差值 $D_1,D_2,\cdots,D_n$ 的绝对值,同时保留每个差值的原有符号。然后,从 $1\sim n$ 分配给这些绝对值相应的整数秩。检验统计量 t 是原来为正的差值的秩之和。大的 t 值表示 X 的总体均值大于 Y 的总体均值,即,要求拒绝 H_0 假设,代之以正的备择假设 $H_1:\mu_D>0$。或者,可计算等于或大于测得的 t 值的右尾概率的 p 值【p Value】。可用于小样本的 t 的精确虚无假设分布表和/或临界值【Critical Value】(Gibbons,1993,1997)。就大样本而言,可以用基于标准正态分布【Normal Distribution】的近似检验,其检验统计量如下,它是 t 的函数,同时加入了一个连续的校正项以改进精度:

$$Z=\frac{T-0.5-[n(n+1)/4]}{\sqrt{n(n+1)(2n+1)/24}}$$

忽略任一等于零的差值,相应地减少 n 的数值。如果任一差值的绝对值相等,那么分配给所有同数值的绝对差值相同的中位-秩,即假定这些绝对差值不相等时,它们可能取得的秩的平均数。

例如,假定有10个驾车人构成的随机样本被选出,用以检验酒精是否对反应时间有逆效应。分别测量驾驶员饮用2盎司80度的酒精饮品前和后的反应时间。数据及反应的秒数显示在表1中。如果酒精对反应时间有逆效应,样本的差值:饮后-饮前,将倾向于有大且正的值。这些差值,它们的绝对值以及秩也在表1中给出。

表 1 Wilcoxon 检验统计量的计算

被试对象	饮前	饮后	差值	绝对值	秩	正差值的秩	负差值的秩
1	0.68	0.73	0.05	0.05	5.5	5.5	
2	0.64	0.63	−0.01	0.01	1.0		−1.0
3	0.68	0.66	−0.02	0.02	2.0		−2.0
4	0.82	0.92	0.10	0.10	9.5	9.5	
5	0.58	0.68	0.10	0.10	9.5	9.5	
6	0.80	0.87	0.07	0.07	8.0	8.0	
7	0.72	0.77	0.05	0.05	5.5	5.5	
8	0.66	0.70	0.04	0.04	3.5	3.5	
9	0.88	0.84	−0.04	0.04	3.5		−3.5
10	0.73	0.79	0.06	0.06	7.0	7.0	
						48.5	

样本 $n=10$ 的检验统计量的值为 $t=48.5$。根据吉布森(Gibbons,1993,p.68),其精确的右尾 p 值为 0.016 5。大样本近似给出 $Z=2.08$,其左尾 p 值为 0.018 8。两个检验的结论都是拒绝虚无假设,并论断酒精对反应时间有逆效应。

在满足假定差值的总体分布是对称且连续的条件下,Wilcoxon 检验才是有效的。当样本量不少于 10 时,正态近似才相当精确,同时,通过在 z 统计量中加入同数值的修正项,可以增进近似的精度。对正态分布而言,该检验几乎同学生 t 检验同样有效,而对其他分布,则可能更有效。许多计算机统计检验软件都包含这一检验。

一个不同的检验,也常称作 Wilcoxon 检验,其实是 Wilcoxon 秩和检验,用以检验两个相互独立的随机样本中等均值的虚无假设,而等价于所谓的 Mann-Whitney *U* 检验【Mann-Whitney *U*-Test】。

——Jean D. Gibbons

(赵锋译校)

参考文献

Gibbons, J. D. (1993). *Nonparametric statistics: An introduction.* Newbury Park, CA: Sage.

Gibbons, J. D. (1997). *Nonparametric methods for quantitative analysis* (3rd ed.). Columbus, OH: American Sciences Press.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, *1*, 80-83.

WinMAX

WinMAX 是一个用于定性资料分析的计算机软件包。它能够对文本资料作扎根理论取向的定性分析,也能做定量类型的内容分析【Content Analysis】。欲获得更多相关信息,请参见该软件的相关网址。

——Tim Futing Liao

(王玥译 赵锋校)

样本内平方和(Within-Samples Sum of Squares)

方差分析【Analysis of Variance】常用于确定不同组间,在某个变量上是否存在差异。这一问题也可以表述成是否存在一个定类变量【Nominal Variable】对一个定量变

量【Quantitative Variable】的效应。如果组均值不同，这必定可归之于某个变量的效应。

理解方差分析的关键在于这一事实，即因变量上的观测值相互区别。如果观测值都相同，那么就不可能有任意变量的效应。如果类别的自变量是唯一对因变量【Dependent Variable】有效应的变量，那么每个组内的观测值就当一致。假如性别是影响收入的唯一变量，那么所有的女性当有相同的收入，所有的男性也当有相同的收入，而且当不同于女性的收入。可以简要地说，所有女性当有同女性收入均值相等的收入，同时，所有男性当有同男性收入均值相等的收入。

但是多数情况下，因变量的观测值是互相不同的，而且也不等于组均值。此一事实，即不是所有的女性都有相同的收入，意味着存在除性别之外的一个或多个变量也对收入有效应。因此，某个组内的观测值的变差【Variation】告诉我们数据中存在其他变量的效应。此外，变差的量级告诉我们这些效应是大还是小。

接着，我们怎样才能测量组内观测值的变差的量级呢？一个可行的做法是看每个观测值同组均值的差异有多大，平方观测值同组均值的差，再总和这些平方。例如，如果某个特定组的观测值是7，9，10和14，其均值为10。组均值的离差为-3，-1，0和4，这些数的平方为9，1，0和16，于是，这些平方的和等于26。这就是用平方和量度某个特定组内观测值的变差的量级。

接着，可以发现每一个组内变差的量级，并将所有组的这些平方和累加。这一累加的和被称为样本内平方和。它也被称为组内平方和，或者简单地称为内平方和或误差平方和。因为组内变差可归之于所有其他的，不同于确定组别的类别变量的变量，所以这一平方和还可以称为残差平方和。使用平方和而非其他方法的理由既有历史的原因，也有数学的原因。

——Gudmund R. Iversen

（赵锋译校）

参考文献

Cobb, G. W. (1998). *Introduction to design and analysis of experiments*. New York: Springer-Verlag.

Iversen, G. R., & Norpoth, H. (1987). *Analysis of variance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, series 07-001, 2nd ed.). Beverly Hills, CA: Sage.

Moore, D. S., & McCabe, G. P. (1998). *Introduction to the practice of statistics*. New York: W. H. Freeman.

受试者内设计（Within-Subject Design）

在统计软件普及之前，人们以易于心算的方式发展实验设计【Experimental Design】。由于确定的公式难以使用，人们便努力简化必要的计算。但是这让我们难以发现设计之间的相似点。受试者内分析非常像重复测量设计【Repeated Measures Design】设计。这两种分析其最简单的样式都可以当作每格仅有一个观测值的双向方差分析【Analysis of Variance】予以处理。

受试者内设计用于在不同的条件下，从几个不同的受试者那里获得相应的测量值，如测验分数。问题是这些测量值间是否存在变化。而变化意味着不同受试者得分的方差更多地应归之于不同的条件，而不是预

期的随机变化。

首先找到某一受试者的均值,再用该受试者的每一测量值减去它的均值,取每一差值的平方,再累加平方,就可以计算出该受试者在测量值上的总方差。由此给出了某一特定受试者在得分上的变异度,也是这一设计名称的由来。在计算出每一受试者的内平方和后,我们累加所有受试者的平方和【Sum of Squares】得到所有的受试者内方差。以符号表示,受试者内平方和等于 $\sum\sum(y_{ij}-\bar{y}_i)^2$,其中 y_{ij}是第 i 个主体的第 j 个测量值,$\bar{y}_i$ 是第 i 个主体的均值。这一总和包含了所有的条件和所有的受试者。若每一受试者有 k 个测量,那么某一受试者的内平方和有 $k-1$ 个自由度【Degrees of Freedom】。若 n 个个体上有数据,那么上述总的受试者内平方和有 $n(k-1)$ 个自由度。

受试者间方差有两个原因:一个可归之于不同条件的效应;另一个可归之于残差变量。

第 i 个主体的第 j 个测量的残差变量的值由下列公式计算得出:

$$(y_{ij}-\bar{y})-(\bar{y}_{i.}-\bar{y})-(\bar{y}_{.j}-\bar{y})=$$
$$y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}$$

从第 i 个受试者的第 j 个测量对总均值的离差中,减去受试者的效应(受试者均值减去总均值),再减去特定条件的效应(条件均值减去总均值)。在去除这两个效应后,剩下的就是残差变量的效应(我们承认,最后的表达式等同于双向方差分析中的交互效应,而当每格只有一个观测值时,则被视作残差效应)。残差平方和有 $(n-1)(k-1)$ 个自由度。

现在,不同条件的平方和可以由受试者内平方和与残差间的差值计算出。它的自由度等于 $n(k-1)-(n-1)(k-1)=k-1$。

如果条件平方和的值较大,那么就存在着由一个条件到下一个条件下测量值上的变化。可通过比较它的均平方和与残差均平方和估得相应的量,其结果是 F 变量上的一个值,以及一个 p 值。

——Gudmund R. Iversen

(王玥译 赵锋校)

参考文献

Cobb, G. W. (1998). *Introduction to design and analysis of experiments*. New York: Springer-Verlag.

Sheskin, D. J. (1997). *Handbook of parametric and nonparametric statistical procedures*. Boca Raton, FL: CRC Press.

Wilcox, R. R. (1996). *Statistics for the social sciences*. San Diego, CA: Academic Press.

书写(Writing)

像所有的语言一样,书写是一种建构的力量,它能创造现实的特定视角,铭记特定的价值。书写的风格可以反映历史的转换范式;社会科学的书写,如所有其他书写,是一种社会历史的建构,既非一成不变,也非客观公正。社会科学中的书写是一个竞技场。一个人如何写、写什么以及为谁写,都成为理论的、伦理的以及方法论的议题,有时被称为"呈现危机"(Clifford & Marcus, 1986)。

历史发展

17 世纪伊始,西方世界的书写被分入两种类型:文学的和科学的。文学作品与虚构、修辞、主观性以及一种含混的(诠释性

的)声音联系在一起,而科学与事实、真理、客观性以及一种明晰的声音联系在一起。整个18世纪,对文学书写的攻击愈加严苛。杰里米·边沁提出,理想的语言应当没有文字,只有符号;大卫·休姆把诗人描绘成职业骗子;塞缪尔·约翰逊试图通过他的字典永久地固定词义。走入这一语言世界,孔多塞引入社会科学一词,声称如果一种关于道德和社会议题的精确语言被采纳,那么错误的想法和行动将几乎不再可能。直至19世纪,文学和科学彼此并立,如两个分离的领域。文学与"艺术"和"文化"结盟,都包含有品位、审美、伦理、人道和道德的价值,而科学与精确和客观性结盟。因为科学的地位高于文学,所以一些文学作家尝试让文学成为科学的一部分。至19世纪末,"现实主义"统治了科学和小说的书写,后者以奥诺雷·德·巴尔扎克和埃米尔·左拉为先锋。

随着20世纪的发展,社会科学书写和文学书写之间的关系变得更加复杂。"现实/虚构""主观/客观""真实/想象"间的界限模糊了。"新报道风格(New Journalists)"让他们自己成为他们书写的主体;英语系把少数民族和后殖民地的小说列为"民族志小说";陷于"呈现危机"的社会科学家,开始用文学形式发表他们的研究,包括诗歌、故事、戏剧(参见创作式分析性实践民族志【Creative Analytical Practice Ethonography】)。1970年代,两种写作的交叉产生了对饰的(oxymoronic)类型,仅举几例,"创造性非小说(creative nonfiction)""纪实文学(faction)""民族志戏剧(ethnographic drama)"和"真的小说(true novel)"。而在1980年代,范梅南(Van Maanen,1988)辨识出三种类型的民族志故事【Ethnographic Tales】——实在论、自白的和印象派的。直至1990年代,在社会科学内部,两类人之间的分裂扩大了,其中一些人坚持科学的书写模式,而另一些人选择借用文学世界的工具,以弥补科学模式的不足。这一分裂没显示出缩小的迹象。

明确选用文学工具的社会科学家中之大部是定性研究者。不像定量研究者用图和表来表现他们的研究发现,定性研究者必须依靠文字。因为当代定性研究的一个目标是:为了读者,从而使其所研究的世界,依其全部的复杂性生动起来。所以,比之于研究论文的惯例和格式,文学工具(如场景设置、对话、多样的立场、风格)和文体(散文、诗歌、信函、对话体等)被证明对此是有帮助的。

尽管大部分定性研究者有意识地选用文学手段,然而还是有一些人继续在社会科学的一般传统下写作。其中许多人把书写简单地看成实际工作或"实地调查"之后的"肃清"活动、"伏案工作"。他们致力于科学模型,同时,其中一些人则借助计算机程序来分析他们的数据。但是,逐渐地,定性研究者把写作本身界定为一种"求知"的方式——不只是"讲"——一种发现、质询和分析的方法。

书写作为质询的方法

作为一种质询方法的书写背离于标准的社会科学实践。它提供了另一种——或备选的——研究实践。作为一种发现方法,书写的目标是去认识某人项目中的新事物、某人自己,以及两者之间的关系。由于心存这一目标,作者探求不同的书写文体、声音、立场,以及风格。由于在某种程度上,类似于与一位可信的朋友进行一场自由自在的对话,一场对其结果没有预设概念的对话,这种书写探索为新的洞见创造了条件——这种洞见在权威文体下是不可能出现的。

作为质询方法的书写与后结构主义【Poststructuralism】的核心是一致的,即怀疑任何一个理论或方法、言语或类型、传统或

新颖都含有对权威性知识的普遍且一般的权利主张。更确切些,后结构主义鼓励地方的、历史的和部分的知识所具有的权利。此外,这种知识随着一个人的生活背景的转换,变换其所依赖的声音、文体、风格,某人写作的此类事项。了解一个人的研究和了解研究者自身是密不可分的,且不可避免地是部分的知识。自我和社会科学是"双生的"。

后结构主义支持三种书写策略。第一,总是出现书面的"我",不管被承认与否,都应该被作者反思性地探求和质询。例如,作者要问,"我为什么要写这一文本? 我的潜在利益是什么?"第二,因为一个人的主体性——书面的"我"——是在一种竞争对话和地方争吵的背景中被造出的,所以一个人的主体性不是固定的、死板的,而在变动着。因此,作者所写的脱离不开作者的生命处境。作者被鼓励,将他们的研究放进不同的框架背景中,如学科的、部门的和家庭的关系(Richardson, 1997)。第三,因为知识是部分的,作者不可能在一个文本中、在一次发声中、在一种文体里说出所有的东西。毋宁,作者被鼓励去从多种立场,并以不同的文体去探求材料,形塑它、剪裁它。多种方式处理民族志材料的例子有:卡罗琳·埃利斯和阿瑟·博克纳主编的《民族志撰写:定性写作的替代形式》(*Composing Ethnography: Alternative Forms of Qualitative Writing*),帕蒂·拉瑟和克里斯汀·史密斯的《困扰天使:携带艾滋病毒或患有艾滋病的女性》(*Troubling the Angels: Women Living With HIV/AIDS*),玛丽·兰格维的《阅读奥斯威辛》(*Reading Auschwitz*),约翰·范梅南的《田野故事说:论民族志写作》(*Tales of the Field: On Writing Ethnography*),劳雷尔·理查森和欧内斯特·洛克里奇的《跟着欧内斯特去旅行:跨越文学/民族志之分野》(*Travels with Ernest: Crossing the Literary-Ethnographic Divide*),沃尔夫的《三重故事:女性主义、后现代和民族志的责任》(*A Thrice Told Tale: Feminism, Postmodernism and Ethnographic Responsibility*;在这本书里,沃尔夫通过一个虚构的故事、她的田野笔记和一篇传统的社会科学文章,讲述了她在台湾地区的研究)。

——Laurel Richardson

(王玥译校)

参考文献

Clifford, J., & Marcus, G. E. (Eds.). (1986). *Writing culture as cultural critique.* Berkeley: University of California Press.

Ellis, C., & Bochner, A. P. (Eds.). (1996). *Composing ethnography: Alternative forms of qualitative writing.* Walnut Creek, CA: Altamira.

Richardson, L. (1997). *Fields of play: Constructing an academic life.* New Brunswick, NJ: Rutgers University Press.

Richardson, L., & Lockridge, E. (in press). *Travels with Ernest: Crossing the literary-ethnographic divide.* Walnut Creek, CA: Altamira.

Van Maanen, J. (1988). *Tales of the field: On writing ethnography.* Chicago: University of Chicago Press.

Wolf, M. (1992). *A thrice told tale: Feminism, postmodernism and ethnographic responsibility.* Stanford, CA: Stanford University Press.

X

变量 X(X Variable)

在回归【Regression】分析中,一个自变量【Independent Variable】会被冠以变量 X,因为它通常以图示的方式被描述成沿某个散点图【Scatterplot】的水平轴[x 轴(x-axis)]变化。有时,它也被称作右侧变量,当它出现在回归方程的右手边时。在某个因果关系中,变量 X 代表假定的因,而变量 Y 则表示果。

设一个标准的回归方程为 $Y=a+bX$。方程右手一侧中的 X 代表一个自变量。思考一下这样一个分析,即城市中人口密度对犯罪率的影响。人口密度就是变量 X,它可能被认为影响犯罪率,即因变量【Dependent Variable】。

相似地,在一个均值差【Difference of Means】检验中,变量 X 是二分变量【Dichotomous Variable】,并被用以计算自变量的均值。例如,如果想要看一下男性和女性在他们的投票参与上是否显著地不同,那么性别就是变量 X。男性和女性的参与率将据此计算,其后,一个均值差【Difference of Means】检验将用于印证该差值是否统计显著。

推广至交叉表【Cross-Tabulations】上,人们通常根据变量 X 内的各个类别计算百分比,并将它们在变量 X 的各类别间作比较,由此来估计自变量对因变量的效应。例如,要比较不同宗教教派的礼拜出席率,教派就是变量 X。定期参加礼拜的天主教徒的比例就可以同相应的新教徒的比例,以及其他宗教定期参加宗教服务的教徒的比例相比较。

——Herbert F. Weisberg

(赵锋译校)

均值 X($\bar{X}$)

$\bar{X}$ 是一个变量【Variable】的算术均值【Mean】的记号法。如果变量记作 X,变量 X 上的第 i 个观测值记作 X_i,且 N 为观测值的总数,那么 $\bar{X}=\sum X_i/N$。图 1 给出了美国 50 个州在小学和初中教育阶段开支的密度图,其均值也标注在图上。

若数据是数值的,均值就最适宜作集中趋势量在【Measures of Central Tendency】,而当变量中有异常值【Outliers】时,中位数【Median】可能更适宜。二分变量【Dichotomous Variables】(计分作"1"或"0")也可以计算其均值。在这种情况下,均值只是类别"1"中个案的比例。尽管存在争论,但是人们也经常计算定序量度

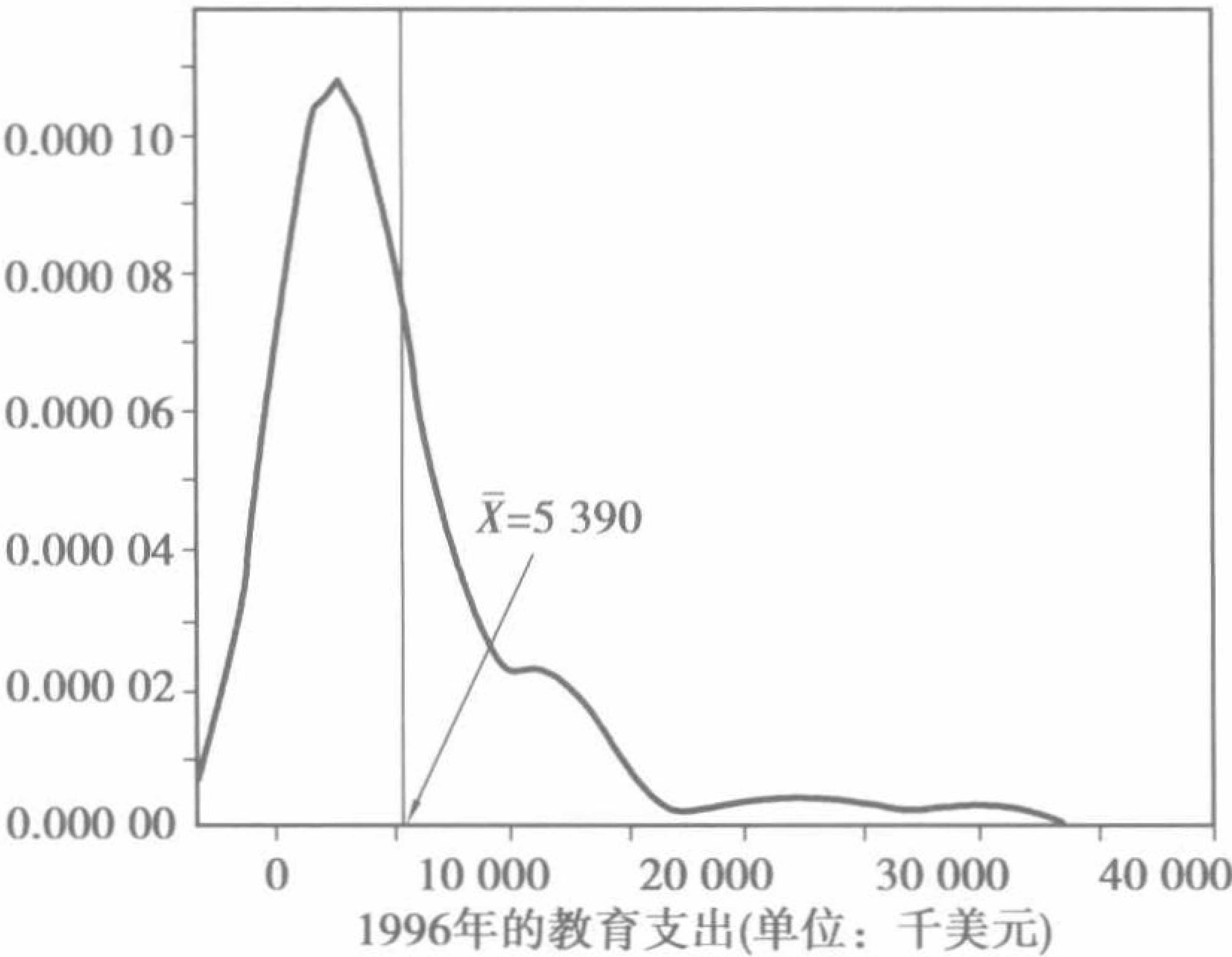


图 1　带均值标注的密度图

资料来源：国家政治学与政策季度数据中库(State Politics & Policy Quarterly Data Resource)。

【Ordinal】的均值。例如，格兰诺威特和沃尔诺(Gravetter & Wallnau, 2000)反对就定序变量计算其均值以及其他传统的统计量，因为观测值间的距离，以及由此而得的均值很难予以明确的定义。然而，博加塔和博恩施泰特(Borgatta & Bohrnstedt, 1980)则坚称在定序变量之下存在着连续变量【Continuous Variables】，所以得到的以整数计分的结果可能相当接近于那些可能得自真正未知的计数类别。

$\overline{X}$ 有几项特定的属性可用于解释均值。第一，围绕均值的计正负号的离差【Deviations】和为零。此即 $\sum(X_i-\overline{X})=0$。此外，围绕任意非均值的计正负号的离差和较大。因此，均值被认为一个"最佳猜测(best guess)"统计量，即在猜测一个观测值的值时，均值将给出比其他猜测值更小的总误差。

第二，由均值计算的正的离差和等于负的离差和(此即围绕均值的计正负号的离差和为零的原因)。根据这一属性，均值有时也被看成变量分布的支点(或平衡点)。在某种程度上某些值高于 $\overline{X}$，相应地就会被某些低于 $\overline{X}$ 的值平衡。

第三，围绕 $\overline{X}$ 的离差平方和小于围绕任意值的离差平方和。均值的最小乘方【Least Squares】属性是根据均值计算方差【Variances】的缘由。

此外，$\overline{X}$ 是变量的总体均值的一个一致的、无偏的【Unbiased】和有效的估计量。它是一致的，因为当观测值的数量接近无穷大时，$\overline{X}$ 减去总体【Population】均值的概率小于任意趋近于 1 的值(δ)。它是无偏的，因为它的期望就是总体均值。同时，因为根据样本【Sample】均值计算的方差要小于由其他总体均值估计量计算的方差，所以它又是有效的。

——David C. Kimball
Herbert F. Weisberg
(赵锋译校)

参考文献

Borgatta, E. F., & Bohrnstedt, G.W. (1980). Level of measurement: Once over again. *Sociological Methods and Research*, *9*, 147-160.

Gravetter, F. J., & Wallnau, L. B. (2000). *Statistics for the behavioral sciences* (5th ed.). Belmont, CA: Wadsworth.

Y

Y 截距(*Y*-Intercept)

在简单回归【Simple Regression】模型中，y 截距是当自变量【Independent Variable】等于零时，因变量【Dependent Variable】的期望值。在回归方程 $Y=a+bX$ 中，a 即 y 截距。图 1 中，在一个带有一条回归线的二维散点图【Scaffer Plot】中，y 截距是回归线同纵轴交叉处的 y 值。在多元回归【Multiple Regression】模型中，截距是当所有自变量等于零时，因变量的期望值。

思考下面的例子，其中研究者有兴趣了解能够减少心脏病或者减少心脏病最致命的后果因素。在特定情况下，研究者希望检查红酒消费(变量 X【X Variable】)和心脏病导致的死亡(变量 Y【Y Variable】)间的关系【Relationship】。主要的假设【Hypothesis】是增加红酒消费可以降低心脏病的风险。研究者从几个国家收集年度红酒消费(每人的升数)和心脏病死亡数(每100 000人口)。由数据估得回归模型为 $Y=261-23X$。在该例中，y-截距为 261，表示在某个假设的没有红酒消费的国家中，模型预测该国每100 000人中有 261 人死于心脏病。下面图 1 给出了相应的数据图，模型的回归线也包含在图中。y 截距的值标注在图上。

因为当 $X=0$ 时，$Y=a$，所以 y 截距也被

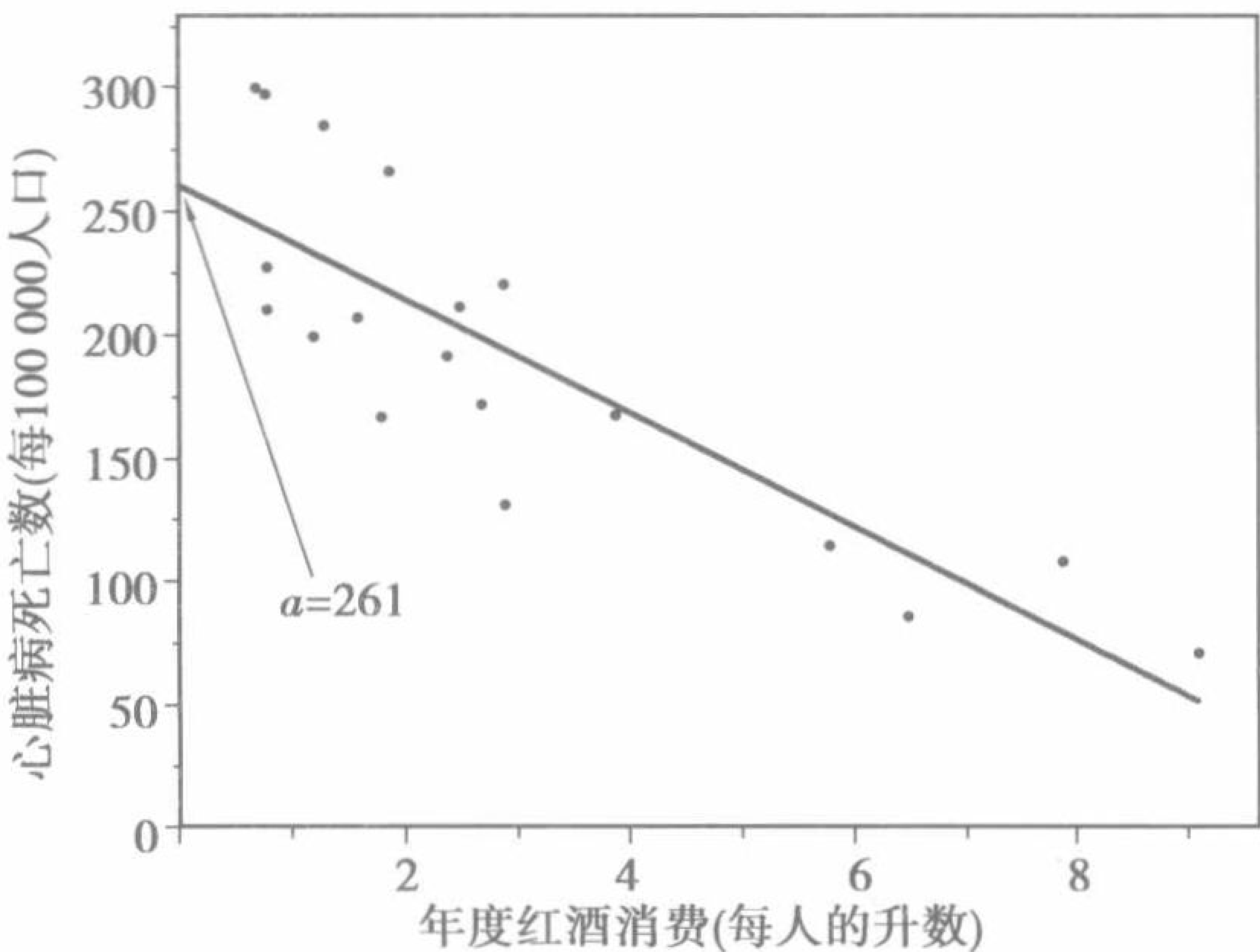


图 1　标有 y-截距的散点图

资料来源：Moore(1998)，p.314。

称为回归方程的"常数项(constant)"。在某些实际情况下，研究者可能出于理论原因，决定自变量为零时，因变量应当始终为零。在这时，可以计算不带常数项(也就是 a 为零)的回归方程，进而简单回归方程变成 $Y=bX$。从图 1 上看，这对应于一条经过原点的回归线。

——David C. Kimball

Herbert F. Weisberg

(赵锋译校)

参考文献

Gujarati, D. N. (1995). *Basic econometrics* (3rd ed.). New York: McGraw-Hill.

Lewis-Beck, M. S. (1995). *Data analysis: An*

introduction. Thousand Oaks, CA: Sage.
Moore, D. S. (1998). *Statistics: Concepts and controversies* (4th ed.). New York: W. H. Freeman.

变量 *Y*(*Y* Variable)

在回归分析中变量 *Y* 是因变量【Dependent Variable】,之所以如此命名是因为它习惯上被描绘为表现两个变量间关系【Relationship】的散点图【Scatterplot】中的纵(*y*)轴。在一个因果关系中,变量 *Y* 代表效应或结果,而变量 *X* 常表示因或自变量。变量 *Y* 通常出现在回归方程的左手一侧;因此,有时又被称为左侧变量。例如,一个简单回归方程常被写成 $Y=a+bX+e$。图 1 给出了一个假想的受教育年限和收入间的关系,其中收入是变量 *Y*,因为我们假设它受到个体受教育年数的影响。

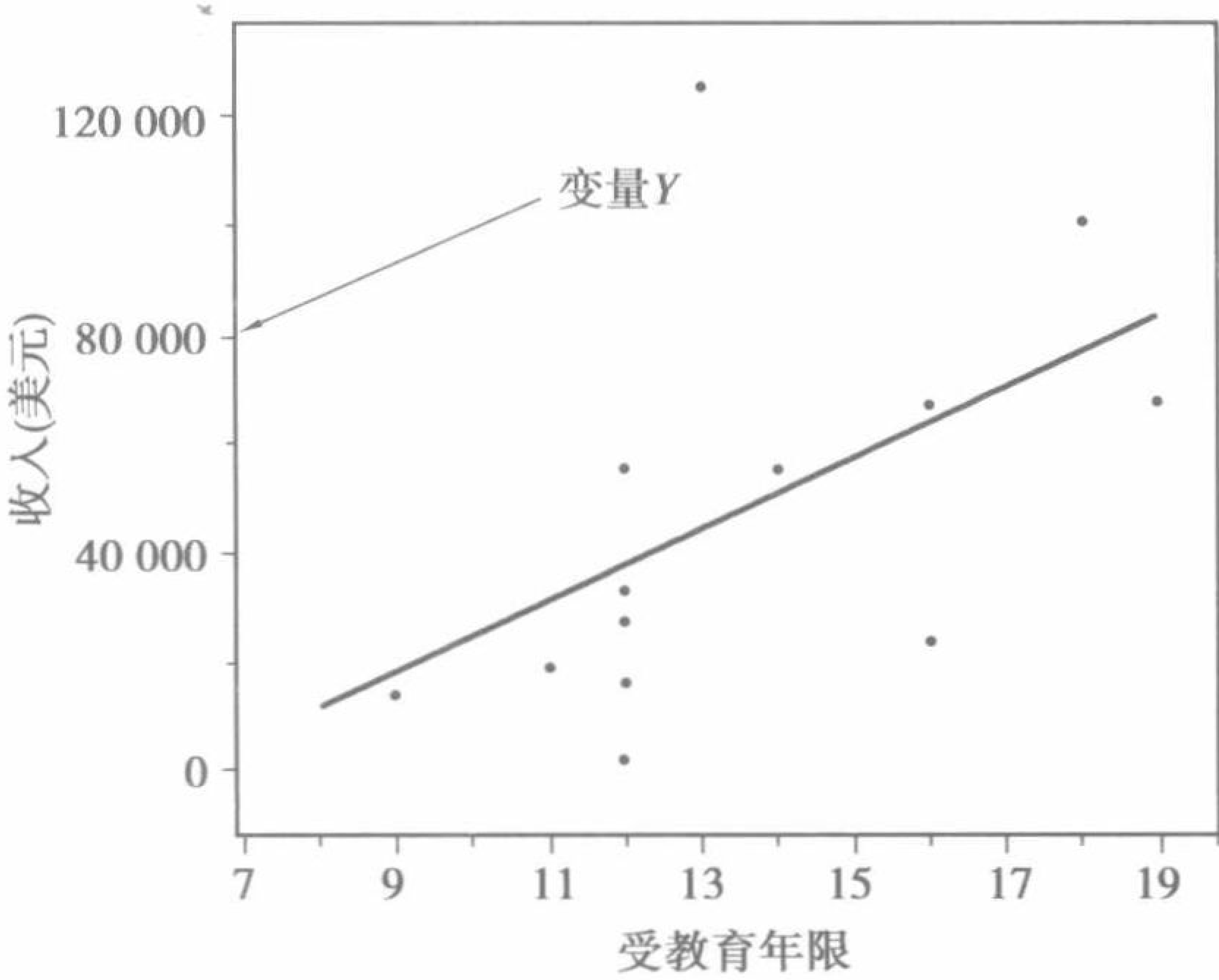


图 1 标示变量 *Y* 的散点图

在两个变量的交叉表【Cross-Tabulations】中,人们通常会计算百分比,以确定一个给定的自变量类别落入因变量的每个类别上的百分比。在均值差【Difference of Means】检验中,变量 *Y* 上的均值【Mean】会根据预测变量【Predictor Variable】的每个类别予以计算,并通过显著性检验【Significance Test】来确定所得到的变量 *Y* 均值间的差值是否大于以机会为基础的期望值。

例如,为了确定不同因素对一辆汽车的油耗的相对贡献,这时,汽车的每加仑英里数就是变量 *Y*,而它的车重、汽缸的数量以及车龄即为合理的预测变量。由此,多元回归模型可以写作 $Y=a+bX+cW+dZ+e$,其中 *Y* 代表因变量(油耗),而 *X*, *W* 和 *Z* 表示预测变量。

——David C. Kimball
Herbert F. Weisberg
(赵锋译校)

Yates 校正(Yates' Correction)

同卡方【Chi-Square】统计量相关的一个校正量。它常用于 2 × 2 的列联表【Contingency Table】中,有时也用于增进带小频次格【Cell】的表的精度。由于认识到该量度可能实际上增加犯第二类错误【Type Ⅱ Error】的概率,现在这一量度已不常用。从根本上说,这一校正会产生较小的卡方值,同时因为它使得统计显著【Statistical Significance】更难确立,所以是对卡方的一个更保守的近似。

——Alan Bryman
(赵锋译校)

Yule Q(Yule's Q)

一个用以表示定类或更高水平的两份变量间关联【Association】强度的量度。Yule Q 也是优势比【Odds Ratio, OR】的函数，其值介于-1～+1。由 OR 得到耶尔 Q 的公式(Walsh & Ollenburger, 2001)如下：

$$Q = \frac{\mathrm{OR} - 1}{\mathrm{OR} + 1}$$

表 1 例示了由一组假设的社会科学数据计算的 Q。在此例中，自变量是性别，因变量是政党。从此数据而得的一项工作假设是男性相对女性更可能成为共和党人。Q 系数可以根据表 1 中格的频次计算。首先以下列方式计算 OR：

$$\mathrm{OR} = \frac{ad}{bc}$$

$$\mathrm{OR} = \frac{10 \times 12}{2 \times 4} = \frac{120}{8} = 15$$

再计算 Q 系数如下：

$$Q = \frac{15 - 1}{15 + 1} = \frac{14}{16} = 0.875$$

表 1　性别对政党选择的关系

	男性	女性
共和党人	$a=10$	$b=2$
民主党人	$c=4$	$d=12$

Q 系数也可根据数据中顺序对与逆序对之差除以顺序对与逆序对之和计算。计算时无须考虑数据中同行对。由此方法计算的数值是顺序对和逆序对的总比值，其数值同由 OR 计算出的结果相同。Q 系数的顺序对/逆序对计算公式(Garson, 2002)如下：

$$Q = \frac{P - Q}{P + Q}$$

当从数据对中的第一个数据转到第二个数据时，自变量和因变量的值都发生变化，且同设定的假设相一致时，数据对是同序的。逆序对也包括两个变量上值的变化，但是数据同假设不一致。同行对有相同的自变量的值，不是计算的组成部分。

$$Q = \frac{ad - bc}{ad + bc}$$

$$Q = \frac{(10 \times 12) - (2 \times 4)}{(10 \times 12) + (2 \times 4)}$$

$$Q = 0.875$$

由这些数据得到的结果显示数据组中的两个变量间有较强的关联。因为在此例中变量是类别的，所以，0.875 的符号是变量排序的结果。颠倒变量的顺序同时会颠倒结果的符号，但是关联强度的量级保持不变。Q 是一个对称的量度，所以选择哪一个变量作为自变量是没有差别的。

Q 可以被解释为减少误差的比例【proportionate reduction of error, PRE】，例如，已知某人的性别可以使我们减少 87.5%的预测因变量的秩的错误。换言之，若已知个体是男性，人们可以猜测他的政党归属为共和党人，而这一猜测很可能是正确的，同时也印证了工作假设。在 Q 等于 0 的情况下，两个变量间的关系是统计独立的，因此，第二个变量(自变量)的知识就不能帮助我们预测第一个变量(因变量)的值。

应用 Q 时存在一个注意事项。理解数据分布而非完全依赖于量度自身是很重要

的一点。如果频次格中的任何一格为零,那么所计算的 Q 表明变量间完全相关(White,1999)。以相同的方式,根据表 2 计算 Q,即

$$Q = \frac{(10 \times 0) - (2 \times 4)}{(10 \times 0) + (2 \times 4)}$$

$$Q = -1.0$$

在这个特别的例子中,-1.0 的完全关联是误导性的,因为在数据组中仅有两个女性回答人。

表 2 性别对政党选择的关系

	男性	女性
共和党人	$a=10$	$b=2$
民主党人	$c=4$	$d=0$

Q 通常也被当作 gamma 系数予以报告,因为对 2 × 2 的列联表而言,它有其特殊的应用。然而,Q 的值通常小于 gamma 的值,因为两分的数据会导致数据损失,进而弱化关联的强度。

——Irvin B. Vann

(赵锋译校)

参考文献

Garson, G. D. (2002, June). *Statnotes: An online textbook* [online]. Retrieved from http://www2.chass.ncsu.edu/ garson/pa765/statnote.htm.

Walsh, A., & Ollenburger, J. C. (2001). *Essential statistics for the social and behavioral sciences: A conceptual approach*. Upper Saddle River, NJ: Prentice Hall.

White, L. (1999). *Political analysis: Technique and practice* (4th ed.). Orlando, FL: Harcourt Brace.

Z

z 分数(z-Score)

z 分数可以通过用变量值减去均值【Mean】，再除以变量的标准差【Standard Deviation】得出。这样一个标准化的分数是一个个案的值离开均值的标准差的数量。它给出了该得分在分布【Distribution】上的相对位置。

在社会科学中，我们常用新的得分代替原来的测量分，即一个被称为变换【Transformation】的方法。原来的得分被称为原始分(raw scores)，新的得分称为变换分。我们可以分别将它们记作 x 和 x'。如果变换是线性的，如 $x'=ax+b$，那么可以肯定无论就均值，还是标准差而言，原始分和变换分之间都有固定的关系。就均值而言，有 $\bar{x}'=a\bar{x}+b$，而对标准差来说，有 $s_{x'}=as_x$。于是有，$a=s_{x'}/s_x$，$b=\bar{x}'-a\bar{x}=\bar{x}'-(s_{x'}/s_x)\bar{x}$。据此，变换分可以写成 $x'=ax+b=(s_{x'}/s_x)x+\bar{x}'-(s_{x'}/s_x)\bar{x}=(s_{x'}/s_x)(x-\bar{x})+\bar{x}'$。在此公式中，原始分的均值和方差($\bar{x}$ 和 s_x)根据数据都是已知的。在变换前，我们决定变换分的均值和标准差的值。一个通行的选择是 $\bar{x}'=0$，$s_{x'}=1$。然后，变换分可写成 $x'=(s_{x'}/s_x)(x-\bar{x})+\bar{x}'=(1/s_x)(x-\bar{x})+0=(x-\bar{x})/s_x$。该变换分被称为标准分或 z 分数，而该变换被称作 z 变换：$z=(x-\bar{x})/s_x$。如果用总体【Population】代替样本，再用希腊字母代替拉丁字母，由此可得 $z=(x-\mu)/\sigma$。

例示

假设你正在研究学生的睡眠行为，同时，你被告知弗朗西斯每晚睡 10 小时。在均值为 8，标准差为 2 的情况下和在均值为 12，标准差为 4 的情况下，这一得分给出的图像是不同的。在第一个分布中，该得分有 $z=(10-8)/2=1$，表示弗朗西斯比均值多睡一个标准差，而在第二个分布中，该得分有 $z=(10-12)/4=-0.5$，表示他比均值少睡半个标准差。因此，在一个分布中，他可能是一个贪睡的人，而在另一个分布中，则略微睡眠不足。

标准化吗?

标准化是有用的，因为它可以给出一组新的、能够加以比较的得分。例如，两个心理试验的得分很难作比。变换到标准分就可以比较了，这样就可以判定一个人在一项试验中的表现是好于还是差于他在另一项试验中的表现。然而，另一方面，标准化作为一种方法，为了获得可比性，它的变量方差被人为地变作了单位 1，同时，人们似乎可以争辩这是一种过激的操作，因为一个个体的得分——比如，弗朗西斯的教育水平——在有着不同教育变差量的社会中有不同的意义。在这种情况下，非标准化的得

分(带有标准差的)可能更适用。一般而言,在需要一个统计模型或相同的统计模型中的多个变量具有可比性时,应当使用标准分,而如果目标是比较用于不同样本的统计模型中的一个或多个变量,则不宜使用标准分。当我们做标准化时,必须报告变量的标准差。

确保可比性的其他方式

最后要说的是,标准化不是为获得不同变量上得分的可比性而变换得分的唯一方式。另一种方式是变换成百分位数【Percentile】。

还应注意的是,在上面给出的所有例子中,变量都假定近似地服从一个正态分布【Normal Distribution】。对于偏斜的分布,以及对于具有不同分布形式的变量的比较,z分作为一个表示个体在某一变量分布上相对位置的指标就不再有效。在这种情形下,非正态的(如偏斜的)分布被变换为一个正态分布。这种变换于是称为正态化【Normalization】,而不是线性变换。

——Jacques Tacq

(赵锋译校)

参考文献

Blalock, H. M. (1972). *Social statistics.* Tokyo: McGraw-Hill Kogakusha.

Hays, W. L. (1972). *Statistics for the social sciences.* New York: Holt International Edition.

Tacq, J. J. A. (1997). *Multivariate analysis techniques in social science research: From problem to analysis.* London: Sage.

Van den Ende, H. W. (1971). *Beschrijvende Statistiek voor Gedragswetenschappen.* Amsterdam/Brussel: Agon Elsevier.

z 检验(*z*-Test)

z 检验用 *z* 分数【*z*-Scores】来实施大样本的显著性检验【Significance Testing】。在抽样分布【Sampling Distribution】具有正态概率密度函数形式的情况下,存在着多种显著性检验。由于正态密度函数会被变换成标准正态概率密度函数【Probability Density Function】,因此检验分就变为 *z* 分数。一个简单的例子是关于总体均值,如 IQ 得分的均值检验。若虚无假设【Null Hypothesis】宣称均值为 100(H_0 : $\mu=100$),那么这是一项精确的假设。若虚无假设宣称均值不大于 100 (H_0 : $\mu\leqslant100$),这就是一项不精确的假设。就后一项假设而言,备择假设可能是均值大于 100(H_a : $\mu>100$)。这时,检验即为单尾的,因为我们假定,样本所来自的人群比平均 IQ 为 100 的人群更聪明。若是双尾检验,备择假设可能是 H_a : $\mu\neq100$。将一个不精确的备择假设同精确的虚无假设结合在一起也是常见的。例如,用备择假设,某个总体的 IQ 均值大于 100(H_a : $\mu>100$)来检验均值等于 100(H_0 : $\mu=100$)的虚无假设。

单个均值的 *z* 检验

当我们想要检验前述总体 IQ 的均值假设时,通过评估比值 $(\bar{x}-\mu)/\sigma_{\bar{x}}$ 可以实施 *Z* 检验——其中 $\bar{x}$ 是样本均值【Mean】,μ 是虚无假设 H_0 下的总体均值,而 $\sigma_{\bar{x}}$ 是均值的标准误【Standard Error】(虚无假设 H_0 下均值的抽样分布的标准差【Standard Deviation】)。这一比值将一个均值差变换成 *z* 分数。现在假设在一个 $n=900$ 的样本(有回置地随机抽取的)中,$\bar{x}=102$, $s=10$。

样本充分地大，并根据中心极限定理【Central Limit Theorem】，抽样分布呈正态分布，其均值等于总体均值 μ，其方差等于 $\sigma_{\bar{x}}=\sigma^2/n$。由于 σ 未知，抽样分布的方差可以估计作 $\hat{\sigma}_{\bar{x}}^2=s^2/n$（以替代 σ^2/n），其平方根记作 $\hat{\sigma}_{\bar{x}}$。我们的样本统计量 $\bar{x}=102$，可以被置于均值为 100、方差为 $s^2/n=10^2/900=0.111$ 的抽样分布之中。于是，它的 z 分数为 $z=(102-100)/\sqrt{0.111}=6$。针对该 z 分数值或更高的值而言，我们在标准正态表中可以查到其概率为 0.000 001，而这个概率是非常小的（远远小于作出第一类错误【Type Ⅰ Error】的概率，即 $\alpha=0.05$，比 $\alpha=0.001$ 还小），所以在虚无假设成立的情况下，发现此样本是非常不可能的。由此，虚无假设被拒，同时，我们有经验证据支持 $\mu>100$ 的备择假设。还需要指出的是，我们不知道犯第二类错误【Type Ⅱ Error】β 的概率有多大，但是在大样本量（$n=900$）的情况下，我们可以有把握认为 β 会很小。

z 检验的其他应用

z 检验还有许多其他的应用。其中之一是比率的检验。以总统选取为例，决定胜选的比率通常是 $\pi=0.50$。所以，我们可以检验这样一个假设，某个候选人有超过半数的选票，即 H_a：$\pi>0.50$。这将与虚无假设 H_0：$\pi=0.50$ 相对。现在，总体的比率记作希腊字母 π，而样本的比率用拉丁字母 p 表示。给定虚无假设为真的情况下，现在，抽样分布就是比率的分布（而非均值的）。这时，该抽样分布具有二项分布【Binomial】的形式。对于大样本，可以再次证明可由正态分布【Normal Distribution】予以替代，因为在大样本的情况下，二项分布的概率和正态分布的概率会趋于一致。比率的抽样分布的均值为总体比率 π（在虚无假设的情况下），而它的方差为 $\pi(1-\pi)/n$，其中 n 为样本量。现在，假设在选取前，以有置回的方式随机抽取，得到一个 $n=900$ 的样本，而且我们发现有 $p=53\%$ 的选票投给了该候选人。给定大样本的情况下，比率的抽样分布具有均值为 $\pi=0.50$，方差为 $\pi(1-\pi)/n=(0.50)(0.50)/900=0.000\ 277\ 7$ 的正态分布形式。由此，可得 $z=(p-\pi)/\sqrt{\pi(1-\pi)/n}=(0.53-0.50)/\sqrt{0.000\ 277\ 7}=1.8$。对于这样一个 z 值或更高的 z 值，在标准正态表中，我们发现其概率为 0.035 9，小于 $\alpha=0.05$，所以在虚无假设成立的条件下，我们将不太可能得到这样一个样本结果。据此，以 $1-\alpha=0.95$ 的可能性，虚无假设被拒，同时，我们有经验证据支持备择假设，即该候选人的选票可能显著地高于 50%。

这里必须指出一项需要注意的事项。二项分布，一个离散分布是由连续的正态分布近似的。因此实际上，p 的值是一个区间的中点，同时，一个“用于连续的校正项”还必须给出。在计算 z 分数时，如果 p 恰好大于 π，我们就从 p 中减去 $0.5/n$；如果 p 恰好小于 π，我们则需要给 p 加上 $0.5/n$。就我们的例子来说，样本比率 p 的 z 分数可以计算为 $z=[p-(0.5/n)-\pi]/\sqrt{\pi(1-\pi)/n}=[0.53-(0.5/900)-0.50]/\sqrt{0.000\ 277\ 7}=1.767$，不同于先前得到的值 1.8。由于样本量很大，这一校正相当小。一般而言，在样本量小的情况下，应当使用连续的校正项。

——Jacques Tacq

（赵锋译校）

参考文献

Blalock, H. M. (1972). *Social statistics.* Tokyo: McGraw-Hill Kogakusha.

Hays, W. L. (1972). *Statistics for the social sciences.* New York: Holt International Edition.

Tacq, J. J. A. (1997). *Multivariate analysis techniques in social science research: From problem*

to analysis. London: Sage.
Van den Ende, H. W. (1971). *Beschrijvende Statistiek voor Gedragswetenschappen*. Amsterdam/Brussel: Agon Elsevier.

零阶(Zero-Order)

参见阶【Order】。

参考文献

Abbott, A., & Forrest, J. (1986). Optimal matching methods for historical sequences. *Journal of Interdisciplinary History*, *16*, 471-494.

Abbott, A., & Hrycak, A. (1990). Measuring resemblance in sequence data. *American Journal of Sociology*, *96*, 144-185.

Abdi, H., Dowling, W. J., Valentin, D., Edelman, B., & Posamentier, M. (2002). *Experimental design and research methods.* Unpublished manuscript, University of Texas at Dallas, Program in Cognition and Neuroscience.

Abdi, H., Valentin, D., & Edelman, B. (1999). *Neural networks.* Thousand Oaks, CA: Sage.

Abdi, H., Valentin, D., Edelman, B., & O'Toole, A. J. (1996). A Widrow-Hoff learning rule for a generalization of the linear auto-associator. *Journal of Mathematical Psychology*, *40*, 175-182.

Abell, P. (1987). *The syntax of social life: The theory and method of comparative narratives.* Oxford, UK: Clarendon.

Abraham, B., & Ledolter, J. (1983). *Statistical methods for forecasting.* New York: Wiley.

Abraham, B., & Ledolter, J. (1986). Forecast functions implied by ARIMA models and other related forecast procedures. *International Statistical Review*, *54*, 51-66.

Achen, C. H. (1975). Mass political attitudes and the survey response. *American Political Science Review*, *69*, 1218-1231.

Achen, C. H. (1977). Measuring representation: Perils of the correlation coefficient. *American Journal of Political Science*, *21*, 805-815.

Achen, C. H. (1982). *Interpreting and using regression.* Beverly Hills, CA: Sage.

Achen, C., & Shively, W. P. (1995). *Cross-level inference.* Chicago: University of Chicago Press.

Achinstein, P. (1983). *The nature of explanation.* New York: Oxford University Press.

Achinstein, P., & Barker, S. F. (1969). *The legacy of logical positivism.* Baltimore: Johns Hopkins University Press.

Adair, J. G., Sharpe, D., & Huynh, C. L. (1989). Hawthorne control procedures in educational experiments: A reconsideration of their use and effectiveness. *Review of Educational Research*, *59*, 215-228.

Aday, L. A. (1996). *Designing and conducting health surveys: A comprehensive guide* (2nd ed.). San Francisco: Jossey-Bass.

Adler, P. A. (1985). *Wheeling and dealing.* New York: Columbia University Press.

Adler, P. A., & Adler, P. (1991). *Backboards and blackboards.* New York: Columbia University Press.

Adler, P. A., & Adler, P. (1994). Observational techniques. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 377-392). Thousand Oaks, CA: Sage.

Adler, P. A., & Adler, P. (1998). *Peer power.* New Brunswick, NJ: Rutgers University Press.

Adler, P., & Adler, P. A. (1987). *Membership roles in field research* (Qualitative Research Methods Series vol. 6). Newbury Park, CA: Sage.

Adorno, T., & Horkheimer, M. (1979). *Dialectic of enlightenment* (J. Cumming, Trans.). London: Verso. (Original work published 1944).

Afifi, A. A., & Clark, V. A. (1996). *Computer-aided multivariate analysis* (3rd ed.). Boca Raton, FL: Chapman & Hall.

Agar, M. (1973). *Ripping and running.* New York: Academic Press.

Aggresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences* (3rd ed.). Upper

Saddle River, NJ: Prentice Hall.

Agresti, A. (1984). *An introduction to categorical data analysis.* New York: John Wiley.

Agresti, A. (1984). *Analysis of ordinal categorical data.* New York: John Wiley.

Agresti, A. (1990). *Categorical data analysis.* New York: John Wiley.

Agresti, A. (1996). *An introduction to categorical data analysis* (2nd ed.). New York: John Wiley.

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley Interscience.

Agresti, A., & Cato, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *American Statistician*, *54*(4), 280-288.

Agresti, A., & Finlay, B. (1997). *Statistical methods for social sciences* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Albrow, M. (1990). *Max Weber's construction of social theory.* London: Macmillan.

Aldrich, J. H., & Cnudde, C. F. (1975). Probing the bounds of conventional wisdom: A comparison of regression, probit, and discriminant analysis. *American Journal of Political Science*, *19*, 571-608.

Aldrich, J. H., & Nelson, F. D. (1984). *Linear probability, logit and probit models.* Beverly Hills, CA: Sage.

Algina, J., Oshina, T. C., & Lin, W.-Y. (1994). Type I error rates for Welch's test and James's second order test under non-normality and inequality of variance when there are two groups. *Journal of Educational and Behavorial Statistics*, *19*, 275-291.

Alkin, M. (1990). *Debates on evaluation.* Newbury Park, CA: Sage.

Allison, P. D. (2003). Event history analysis. In M. A. Hardy & A. Bryman (Eds.), *Handbook of data analysis.* London: Sage.

Allison, P. D. (1978). Measures of inequality. *American Sociological Review*, *43*, 865-880.

Allison, P. D. (1987). Estimation of linear models with incomplete data. In C. C. Clogg (Ed.), *Sociological methodology 1987.* San Francisco: Jossey-Bass.

Allison, P. D. (1999). *Multiple regression: A primer.* Thousand Oaks, CA: Pine Forge.

Allison, P. D. (2001). *Missing data.* Thousand Oaks, CA: Sage.

Allport, G. (1942). *The use of personal documents in psychological science.* New York: Social Science Research Council.

Alonso, E. (2002). AI and agents: State of the art. *AI Magazine*, *23*(3), 25-29.

Altheide, D. L. (1987). Ethnographic content analysis. *Qualitative Sociology*, *10*, 65-77.

Altheide, D. L. (1996). *Qualitative media analysis.* Newbury Park, CA: Sage.

Altheide, D. L. (2002). *Creating fear: News and the construction of crisis.* Hawthorne, NY: Aldine de Gruyter.

Altheide, D., & Johnson, J. (1994). Criteria for assessing interpretive validity in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 485-499). Thousand Oaks, CA: Sage.

Altman, D. G., Machin, D., Bryant, T. N., & Gardner, M. J. (2000). *Statistics with confidence: Confidence intervals and statistical guidelines* (2nd ed.). London: British Medical Journal Books.

Altman, M., Gill, J., & McDonald, M. P. (2003). *Numerical methods in statistical computing for the social sciences.* New York: Wiley.

Alvarez, R. M., & Glasgow, G. (2000). Two-stage estimation of nonrecursive choice models. *Political Analysis*, *8*, 147-165.

Alveson, M., & Sköldberg, K. (2000). *Reflexive methodology.* London: Sage.

Alvesson, M. (2002). *Postmodernism and social research.* Buckingham, UK: Open University Press.

Alvesson, M., & Deetz, S. (2000). *Doing critical management research.* London: Sage.

Amemiya, T. (1985). *Advanced econometrics.* Cambridge, MA: Harvard University Press.

Amemiya, T. (1994). *Introduction to statistics and econometrics.* Cambridge, MA: Harvard University Press.

American Association for Public Opinion Research (AAPOR). (2000). *Standard definitions: Final*

dispositions for case codes and outcome rates for surveys. Ann Arbor, MI: Author.

American Educational Research Association, American Psychological Association, & National Council on Educational Measurement. (1999). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. *American Psychologist*, *47*, 1597-1611.

American Psychological Association. (2002, October 28). *PsycINFO database records* [electronic database]. Washington, DC: Author.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Anastasi, A., & Urbina, S. (1996). *Psychological testing* (7th ed.). New York: Prentice Hall.

Anderberg, M. R. (1973). *Cluster analysis for applications.* New York: Academic Press.

Anderson, N. H. (1981). *Methods of information integration theory.* New York: Academic Press.

Anderson, S., Auquier, A., Hauck, W. W., Oakes, D., Vandaele, W., & Weisberg, H. I. (1980). On the use of matrices in certain population mathematics. In *Statistical methods for comparative studies: Techniques for bias reduction.* New York: Wiley.

Andersson, B. E., & Nilsson, S. G. (1964). Studies in the reliability and validity of the critical incident technique. *Journal of Applied Psychology*, *48*(1), 398-403.

Andrews, F. M., Morgan, J. N., & Sonquist, J. A. (1967). *Multiple classification analysis: A report on a computer program for multiple regression using categorical predictors.* Ann Arbor, MI: Survey Research Center, Institute for Social Research.

Andrews, L., & Nelkin, D. (1998). Do the dead have interests? Policy issues of research after life. *American Journal of Law & Medicine*, *24*, 261.

Angle, J. (1986). The Surplus Theory of Social Stratification and the size distribution of personal wealth. *Social Forces*, *65*, 293-326.

Angle, J. (2002). The statistical signature of pervasive competition on wage and salary incomes. *Journal of Mathematical Sociology*, *26*, 217-270.

Angrasino, M. V., & Mays de Perez, K. (2000). Rethinking observation: From method to context. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 673-702). Thousand Oaks, CA: Sage.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*, 444-455.

Angst, D. B., & Deatrick, J. A. (1996). Involvement in health care decisions: Parents and children with chronic illness. *Journal of Family Nursing*, *2*(2), 174-194.

Anselin, L. (1988). *Spatial econometrics.* Boston: Kluwer Academic.

Anselin, L., & Bera, A. (1998). Spatial dependence in linear regression models, with an introduction to spatial econometrics. In A. Ullah & D. Giles (Eds.), *Handbook of applied economic statistics* (pp. 237-289). New York: Marcel Dekker.

Antoni, M. H., Baggett, L., Ironson, G., LaPerriere, A., August, S., Kilmas, N., et al. (1991). Cognitive-behavioral stress management intervention buffers distress responses and immunologic changes following notification of HIV-1 seropositivity. *Journal of Consulting and Clinical Psychology*, *59*(6), 906-915.

Arabie, P., & Boorman, S. A. (1973). Multidimensional scaling of measures of distance between partitions. *Journal of Mathematical Psychology*, *10*, 148-203.

Arbuckle, J. (1997). *Amos users' guide version 3.6.* Chicago: Smallwaters Corporation.

Arch, D. C., Bettman, J. R., & Kakkar, P. (1978). Subjects' information processing in information display board studies. In K. Hunt (Ed.), *Advances in consumer research*, *Vol. 5* (pp. 555-559). Ann Arbor, MI: Association for Consumer Research.

Archer, M. (1995). *Realist social theory: The morphogenetic approach.* Cambridge, UK:

Cambridge University Press.

Archer, M., Bhaskar, R., Collier, A., Lawson, T., & Norrie, A. (Eds.). (1998). *Critical realism: Essential readings.* London: Routledge Kegan Paul.

Aristotle. (1977). *Metaphysica.* Baarn: Het Wereldvenster.

Armacost, R. L., Hosseini, J. C., Morris, S. A., & Rehbein, K. A. (1991). An empirical comparison of direct questioning, scenario, and randomized response methods for obtaining sensitive business information. *Decision Sciences*, *22*, 1073-1090.

Armitage, P., & Colton, T. (Eds.). (1998). *Encyclopedia of biostatistics.* New York: John Wiley.

Armstrong, J. S. (Ed.). (2001). *Principles of forecasting.* Boston: Kluwer.

Aronow, E., Reznikoff, M., & Moreland, K. (1994). *The Rorschach technique.* Needham Heights, MA: Allyn & Bacon.

Aronson, E., Carlsmith, M., &Ellsworth, P. C. (1990). *Methods of research in social psychology.* New York: McGraw-Hill.

Arrow, K. J. (1963). *Social choice and individual values* (2nd ed.). New Haven, CT: Yale University Press.

Arthur, W. B. (1994). *Increasing returns and path dependence in the economy.* Ann Arbor: University of Michigan Press.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology*, *41*, 258-290.

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership, and men* (pp. 177-190). Pittsburgh, PA: Carnegie Press.

Asher, H. B. (1983). *Causal modeling* (2nd ed., Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-003). Beverly Hills, CA: Sage.

Ashmore, M. (1989). *The reflexive thesis: Wrighting sociology of scientific knowledge.* Chicago: University of Chicago Press.

Ashmore, M. (1995). Fraud by numbers: Quantitative rhetoric in the Piltdown forgery discovery. *South Atlantic Quarterly*, *94*(2), 591-618.

Atkinson, J. M., & Heritage, J. (Eds.). (1984). *Structures of social action: Studies in conversation analysis.* Cambridge, UK: Cambridge University Press.

Atkinson, P. (1990). *The ethnographic ima gination: Textual constructions of reality.* London: Routledge.

Atkinson, P., & Hammersley, M. (1994). Ethnography and participant observation. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 248-261). Thousand Oaks, CA: Sage.

Atkinson, P., & Silverman, D. (1997). Kundera's "immortality": The interview society and the invention of self. *Qualitative Inquiry*, *3*(3), 304-325.

Atkinson, P., Coffey, A., & Delamont, S. (2001). Editorial: A debate about our canon. *Qualitative Research*, *1*, 5-21.

Atkinson, R. (1998). *The life story interview* (Qualitative Research Methods Series, Vol. 44). Thousand Oaks, CA: Sage.

Axelrod, R. (1984). *The evolution of cooperation.* New York: Basic Books.

Axelrod, R. (1995). A model of the emergence of new political actors. In N. Gilbert & R. Conte (Eds.), *Artificial societies.* London: UCL.

Axelrod, R. (1997). *The complexity of cooperation: Agentbased models of competition and collaboration.* Princeton, NJ: Princeton University Press.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*, 1390-1396.

Ayer, A. J. (1936). *Language, truth and logic.* London: Victor Gollancz.

Babbie, E. (1979). *The practice of social research.* Belmont, CA: Wadsworth.

Babbie, E. R. (1995). *The practice of social research* (7th ed.). Belmont, CA: Wadsworth.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, *73*, 5-59.

Bainbridge, E. E., Carley, K. M., Heise, D. R., Macy, M. W., Markovsky, B., & Skvoretz, J. (1994). Artificial social intelligence. *Annual Review of Sociology*, *20*, 407-436.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis*

(2nd ed.). New York: Cambridge University Press.

Bakeman, R., & Quera, V. (1995). *Analyzing interaction: Sequential analysis with SDIS and GSEQ*. New York: Cambridge University Press.

Baker, B. O., Hardyck, C. D., & Petrinovich, L. F. (1966). Weak measurement vs. strong statistics: An empirical critique of S.S. Stevens' proscriptions on statistics. *Educational and Psychological Measurement*, *26*, 291-309.

Baltagi B. H. (2001). *Econometric analysis of panel data*. Chichester, UK: Wiley.

Banerjee, A., Dolado, J. J., Galbraith, J., & Hendry, D. F. (1993). *Cointegration, error correction and the econometric analysis of nonstationary series*. New York: Oxford University Press.

Banks, M. (2001). *Visual methods in social research*. London: Sage.

Bargh, J. A., & Chartrand, T. L. (2000). Mind in the middle: A practical guide to priming and automaticity research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 253-285). New York: Cambridge University Press.

Barnard, J., & Rubin, D. B. (1999). Small-sample degrees of freedom with multiple imputation. *Biometrika*, *86*, 948-955.

Barndt, D. (1997). Zooming out/zooming in: Visualizing globalization. *Visual Sociology*, *12*(2), 5-32.

Barnes, D. B., Taylor-Brown, S., & Weiner, L. (1997). "I didn't leave y'all on purpose": HIV-infected mothers' videotaped legacies for their children. *Qualitative Sociology*, *20*(1), 7-32.

Barnett, V. (1974). *Elements of sampling theory*. Sevenoaks, UK: Hodder & Stoughton.

Barr, R. (1996). A comparison of aspects of the US and UK censuses of population. *Transactions in GIS*, *1*, 49-60.

Barthes, R. (1967). *Elements of semiology*. London: Jonathan Cape.

Barthes, R. (1977). *Image music text*. London: Fontana.

Barthes, R. (1977). Rhetoric of the image. In *Image, music, text* (S. Heath, Trans.). London: Fontana.

Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis*. London: Arnold.

Bartlett, M. S. (1947). Multivariate analysis. *Journal of the Royal Statistical Society, Series B*, *9*, 176-197.

Bartunek, J. M., & Louis, M. R. (1996). *Insider-outsider team research*. Thousand Oaks, CA: Sage.

Bates, D. M., & Watts, D. G. (1988). *Nonlinear regression analysis and its applications*. New York: John Wiley.

Bateson, G., & Mead, M. (1942). *Balinese character: A photographic analysis*. New York: New York Graphic Society.

Baudrillard, J. (1998). *The consumer society*. London: Sage. (Originally published in 1970)

Bauman, Z. (1978). *Hermeneutics and social science*. London: Hutchinson.

Baumrind, D. (1964). Some thoughts on ethics after reading Milgram's "Behavioral study of obedience." *American Psychologist*, *19*, 421-423.

Bayer, A. E., & Smart, J. C. (1991). Career publication patterns and collaborative "styles" in American academic science. *Journal of Higher Education*, *62*, 613-636.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, *53*, 370-418.

Bayes, T. (1958). An essay towards solving a problem in the doctrine of chances. *Biometrika*, *45*, 293-315.

Baym, N. (2000). *Tune in, log on: Soaps, fandom and online community*. Thousand Oaks, CA: Sage.

Beck, N. (1991). Comparing dynamic specifications: The case of presidential approval. *Political Analysis*, *3*, 51-87.

Beck, N., & Jackman, S. (1998). Beyond linearity by default: Generalized additive models. *American Journal of Political Science*, *42*, 596-627.

Beck, N., & Katz, J. N. (1995). What to do (and not to do) with time-series cross-section data. *American Political Science Review*, *89*, 634-647.

Becker, H. S. (1967). Whose side are we on? *Social Problems*, *14*, 239-247.

Becker, H. S. (1974). Photography and socioiogy. *Studies in the Anthropology of Visual*

Communication, *1*(1), 3-26.

Becker, H. S. (1986). *Doing things together.* Evanston, IL: Northwestern University Press.

Becker, H. S. (1998). *Tricks of the trade.* Chicago: University of Chicago Press.

Becker, H. S., Hughes, E. C., & Strauss, A. L. (1961). *Boys in white.* Chicago: University of Chicago Press.

Becker, S., & Bryman, A. (Eds.). (2004). *Understanding research for social policy and practice: Themes, methods and approaches.* Bristol, UK: Policy Press.

Bell, C. (1969). A note on participant observation. *Sociology*, *3*, 417-418.

Bell, S. E. (1999). Narratives and lives: Women's health politics and the diagnosis of cancer for DES daughters. *Narrative Inquiry*, *9*(2), 1-43.

Bellman, B. L., & Jules-Rosette, B. (1977). *A paradigm for looking: Cross-cultural research with visual media.* Norwood, NJ: Ablex.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity.* New York: John Wiley.

Beltrami, E. J. (1999). *What is random? Chance and order in mathematics and life.* New York: Copernicus.

Benfer, R. A., Brent, E. E., Jr., & Furbee, L. (1991). *Expert systems.* Newbury Park, CA: Sage.

Bennet, D. J. (1998). *Randomness.* Cambridge, MA: Harvard University Press.

Bennett, A. (1999). *Condemned to repetition? The rise, fall, and reprise of Soviet-Russian military interventionism, 1973-1996.* Cambridge: MIT Press.

Benton, T. (1977). *Philosophical foundations of the three sociologies.* London: Routledge Kegan Paul.

Benton, T. (1998). Realism and social science. In M. Archer, R. Bhaskar, A. Collier, T. Lawson, & A. Norrie (Eds.), *Critical realism: Essential readings.* London: Routledge Kegan Paul.

Benzécri, J.-P. (1973). *Analyse des Données. Tome 2: Analyse des Correspondances* [Data analysis: Vol. 2. Correspondence analysis]. Paris: Dunod.

Berelson, B. (1954). Content analysis. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 1, pp. 488-522). Reading, MA: Addison-Wesley.

Berg, B. L. (2001). *Qualitative research methods for the social sciences.* Boston: Allyn and Bacon.

Berger, J. O., & Wolpert, R. L. (1984). *The likelihood principle: A review, generalizations, and statistical implications.* Hayward, CA: Institute of Mathematical Statistics.

Berger, J., Fisek, H., Norman, R., & Zelditch, M. (1977). *Status characteristics and social interaction: An expectation states approach.* New York: Elsevier.

Berger, P., & Luckmann, T. (1967). *The social construction of reality.* Harmondsworth, UK: Penguin.

Bergner, R. M. (1974). The development and evaluation of a training videotape for the resolution of marital conflict. *Dissertation Abstracts International*, *34*, 3485B. (UMI No. 73-32510)

Bergsma, W. (1997). *Marginal models for categorical data.* Tilburg, The Netherlands: Tilburg University Press.

Berk, R. A., & Freedman, D. A. (1995). Statistical assumptions as empirical commitments. In T. G. Blomberg & S. Cohen (Eds.), *Law, punishment, and social control: Essays in honor of Sheldon Messinger* (pp. 245-258). New York: Aldine de Gruyter.

Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian theory.* New York: Wiley.

Bernstein, R. (1983). *Beyond objectivism and relativism: Science, hermeneutics and praxis.* Oxford, UK: Basil Blackwell.

Berry, J. (2002). Validity and reliability issues in elite interviewing. *PS—Political Science and Politics*, *35*, 679-682.

Berry, W. D. (1984). *Nonrecursive causal models.* Beverly Hills, CA: Sage.

Berry, W. D. (1993). *Understanding regression assumptions.* Newbury Park, CA: Sage.

Beverley, J. (1989). The margin at the center: On testimonio. *Modern Fiction Studies*, *35*(1), 11-28.

Beverley, J., & Zimmerman, M. (1990). *Literature and politics in the Central American revolutions.* Austin: University of Texas Press.

Bhaskar, R. (1978). *A realist theory of science.* Hassocks, UK: Harvester Press.

Bhaskar, R. (1979). *The possibilities of naturalism: A philosophical critique of the contemporary human*

sciences. Brighton, UK: Harvester.

Bhaskar, R. (1986). *Scientific realism and human emancipation*. London: Verso.

Bhaskar, R. (1997). *A realist theory of science*. London: Verso. (Original work published 1978)

Bhaskar, R. (1998). *The possibility of naturalism* (2nd ed.). Hemel Hempstead, UK: Harvester Wheatsheaf. (Original work published 1979)

Bickman, L., & Rog, D. J. (Eds.). (1998). *Handbook of applied social research methods*. Thousand Oaks, CA: Sage.

Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A., & Sudman, S. (Eds.). (1991). *Measurement errors in surveys*. New York: Wiley.

Biggs, D., De Ville, B., & Suen, E. (1991). A method of choosing multiway partitions for classification and decision trees. *Journal of Applied Statistics*, *18*, 49-62.

Bijleveld, C. C. J. H., & van der Kamp, L. J. Th. (with Mooijaart, A., van der Kloot, W. A., van der Leeden, R., & van der Burg, E.). (1998). *Longitudinal data analysis: Designs, models, and methods*. London: Sage.

Billig, M. (1996). *Arguing and thinking*. Cambridge, UK: Cambridge University Press.

Bimler, D., & Kirkland, J. (1999). Capturing images in a net: Perceptual modeling of product descriptors using sorting data. *Marketing Bulletin*, *10*, 11-23.

Birdwhistell, R. L. (1970). *Kinesics and context: Essays on body motion communication*. Philadelphia: University of Pennsylvania Press.

Birnbaum, A. L. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Birnbaum, M. H. (2001). *Introduction to behavioral research on the Internet*. Upper Saddle River, NJ: Prentice Hall.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford, UK: Oxford University Press.

Bishop, G., & Smith, A. (2001). Response-order effects and the early Gallup split-ballots. *Public Opinion Quarterly*, *65*, 479-505.

Black, D. (1958). *The theory of committees and elections*. Cambridge, UK: Cambridge University Press.

Black, T. R. (1999). *Doing quantitative research in the social sciences: An integrated approach to research design, measurement, and statistics*. London: Sage.

Blaikie, N. (1993). *Approaches to social enquiry*. Cambridge, UK: Polity.

Blaikie, N. (2000). *Designing social research: The logic of anticipation*. Cambridge, UK: Polity.

Blalock, H. (1961). *Causal inference in nonexperimental research*. Chapel Hill: University of North Carolina Press.

Blalock, H. M. (1960). *Social statistics*. New York: McGraw-Hill.

Blalock, H. M. (1961). Theory, measurement and replication in the social sciences. *American Journal of Sociology*, *66*(1), 342-347.

Blalock, H. M. (1972). *Social statistics* (2nd ed.). Tokyo: McGraw-Hill Kogakusha.

Blalock, H. M. (1979). *Social statistics* (rev. 2nd ed.). New York: McGraw-Hill.

Blalock, H. M., Jr. (1969). *Theory construction: From verbal to mathematical formulations*. Englewood Cliffs, NJ: Prentice Hall.

Blanck, P. D. (Ed.). (1993). *Interpersonal expectations: Theory, research, and applications*. New York: Cambridge University Press.

Blascovich, J. (2000). Psychophysiological indexes of psychological processes. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 117-137). Cambridge, UK: Cambridge University Press.

Blascovich, J., & Tomaka, J. (1996). The biopsychosocial model of arousal regulation. In M. Zanna (Ed.), *Advances in experimental social psychology* (pp. 1-51). New York: Academic Press.

Blau, P. M. (1974). Parameters of social structure. *American Sociological Review*. *39*, 615-635.

Bloor, M. (1997). Addressing social problems through qualitative research. In D. Silverman (Ed.), *Qualitative research: Theory, method and practice* (pp. 221-238). London: Sage.

Blossfeld, H.-P., & Rohwer, G. (2002). *Techniques*

of event history modeling: New approaches to causal analysis. Mahwah, NJ: Lawrence Erlbaum Associates.

Blumen, I. M., Kogan, M., & McCarthy, P. J. (1955). *The industrial mobility of labor as a probability process.* Ithaca, NY: Cornell University Press.

Blumen, I., Kogan, M., & McCarthy, P. J. (1968). Probability models for mobility. In P. F. Lazarsfeld & N. W. Henry (Eds.), *Readings in mathematical social science* (pp. 318-334). Cambridge: MIT Press.

Blumer, H. (1956). Sociological analysis and the "variable." *American Sociological Review, 21*, 683-690.

Blumer, H. (1969). *An empirical appraisal of Thomas and Znaniecki (1918-20) The Polish Peasant in Europe and America.* New Brunswick, NJ: Transaction Books. (Original work published 1939)

Blumer, H. (1969). *Symbolic interactionism: Perspective and method.* Englewood Cliffs, NJ: Prentice Hall.

Blundell, R.W., & Smith, R. J. (1989). Estimation in a class of simultaneous equation limited dependent variable models. *Review of Economic Studies, 56*, 37-58.

Blurton Jones, N. (Ed.). (1972). *Ethological studies of child behaviour.* London: Cambridge University Press.

Bobko, P. (1995). *Correlation and regression: Principles and applications for industrial organizational psychology and management.* New York: McGraw-Hill.

Bogartz, R. S. (1994). *An introduction to the analysis of variance.* Westport, CT: Praeger.

Bogdewic, S. P. (1999). Participant observation. In B. J. Crabtree & W. L. Miller (Eds.), *Doing qualitative research* (pp. 47-69). Thousand Oaks, CA: Sage.

Bohman, J. (1991). *New philosophy of social science: Problems of indeterminacy.* Cambridge, UK: Polity.

Boje, D. M. (2001). *Narrative methods for organizational and communication research.* Thousand Oaks, CA: Sage.

Bolger, N., Davis, A., & Rafaeli, E. (in press). Diary methods: Capturing life as it is lived. *Annual Review of Psychology.*

Bollen, K. (1989). *Structural equations with latent variables.* New York: John Wiley.

Bollen, K. A., & Curran, P. J. (in press). *Latent curve models: A structural equation approach.* New York: John Wiley.

Bollen, K. A., & Jackman, R. W. (1990). Regression diagnostics: An expository treatment of outliers and influential cases. In J. Fox & J. S. Long (Ed.), *Modern methods of data analysis* (pp. 257-291). Newbury Park, CA: Sage.

Bonnie, R. J., & Wallace, R. B. (2002). *Elder mistreatment: Abuse, neglect, and exploitation in an aging America.* Washington, DC: Joseph Henry Press.

Boomsma, A., Van Duijn, M. A. J., & Snijders, T. A. B. (Eds.). (2001). *Essays on item response theory.* New York: Springer.

Borg, I., & Groenen, P. (1997). *Modern multidimensional scaling.* New York: Springer-Verlag.

Borgatta, E. F., & Bohrnstedt, G.W. (1980). Level of measurement: Once over again. *Sociological Methods and Research, 9*, 147-160.

Borgatta, E. F., & Jackson, D. J. (Eds.). (1980). *Aggregate data: Analysis and interpretation.* Beverly Hills, CA: Sage.

Bornstein, R. F. (2002). A process dissociation approach to objective-projective test score interrelationships. *Journal of Personality Assessment, 78*, 47-68.

Bornstein, R. F., Rossner, S. C., Hill, E. L., & Stepanian, M. L. (1994). Face validity and fakability of objective and projective measures of dependency. *Journal of Personality Assessment, 63*, 363-386.

Boruch, R. F. (1997). *Randomized experiments for planning and evaluation: A practical guide.* Thousand Oaks, CA: Sage.

Boruch, R. F., & Cecil, J. S. (1979). *Assuring the confidentiality of social research data.* Philadelphia: University of Pennsylvania Press.

Boster, J. S. (1994). The successive pile sort. *Cultural Anthropology Methods, 6*(2), 7-8.

Bosworth, M. (1999). *Engendering resistance: Agency and power in women's prisons.* Aldershot, UK:

Ashgate/ Dartmouth.

Bottorff, J. L. (1994). Using videotaped recordings in qualitative research. In J. M. Morse (Ed.), *Critical issues in qualitative research* (pp. 224-261). Newbury Park, CA: Sage.

Boucke, O. F. (1923). The limits of social science: II. *American Journal of Sociology*, *28*(4), 443-460.

Bourdieu, P. (1977). *Outline of a theory of practice.* Cambridge, UK: Cambridge University Press.

Bourdieu, P. (2000). *Pascalian meditations* (R. Nice, Trans.). Cambridge, UK: Polity.

Bourdieu, P., & Waquant, L. J. D. (1992). *An invitation to reflexive sociology.* Chicago: University of Chicago Press.

Bourque, L. B., & Clark, V. A. (1992). *Processing data: The survey example* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-085) Newbury Park, CA: Sage.

Bourque, L. B., & Fielder, E. P. (2003). *How to conduct self-administered and mail surveys* (2nd ed.). Thousand Oaks, CA: Sage.

Bourque, L. B., & Russell, L. A. (with Krauss, G. L., Riopelle, D., Goltz, J. D., Greene, M., McAffee, S., & Nathe, S.) (1994, July). *Experiences during and responses to the Loma Prieta earthquake.* Oakland: Governor' s Office of Emergency Services, State of California.

Bourque, L. B., Shoaf, K. I., & Nguyen, L. H. (1997). Survey research. *International Journal of Mass Emergencies and Disasters*, *15*, 71-101.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society (B)*, *26*(2), 211-252.

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control* (rev. ed.). San Francisco: Holden-Day.

Box, G. E. P., & Tiao, G. C. (1965). A change in level of nonstationary time series. *Biometrika*, *52*, 181-192.

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis.* New York: Wiley.

Box, G. E. P., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, *70*, 70-92.

Box, G. E. P., Hunter, W. G., & Hunter, J. S. (1978). *Statistics for experimenters.* New York: Wiley.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis: Forecasting and control* (3rd ed.). New York: Prentice Hall.

Box-Steffensmeier, J. M., & Jones, B. S. (1997). Time is of the essence: Event history models in political science. *American Journal of Political Science*, *41*, 336-383.

Box-Steffensmeier, J. M., & Zorn, C. J. W. (2001). Duration models and proportional hazards in political science. *American Journal of Political Science*, *45*, 951-967.

Box-Steffensmeier, J. M., & Zorn, C. J. W. (2002). Duration models for repeated events. *The Journal of Politics*, *64*(4), 1069-1094.

Boyd, J. P. (1990). *Social semigroups: A unified theory of scaling and bockmodeling as applied to social networks.* Fairfax, VA: George Mason University Press.

Boyd, J. P., & Jonas, K. J. (2001). Are social equivalences ever regular? Permutation and exact tests. *Social Networks*, *23*, 87-123.

Bracken, B. (Ed.). (1995). *Handbook of self concept: Developmental, social and clinical consequences.* New York: Wiley.

Bradley, J. V. (1985). *Distribution-free statistical tests.* Englewood Cliffs, NJ: Prentice Hall.

Brady, I. (2000). Anthropological poetics. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 949-979). Thousand Oaks, CA: Sage.

Brady, I. (2003). *The time at Darwin' s Reef: Poetic explorations in anthropology and history.* Walnut Creek, CA: AltaMira.

Bray, J. H., & Maxwell, S. E. (1985). *Multivariate analysis of variance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-054) Beverly Hills, CA: Sage.

Breen, R. (1996). *Regression models: Censored, sample selected, or truncated data* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-111). Thousand Oaks, CA: Sage.

Breiman, L., Friedman, J., Olshen, R., & Stone, C.

(1984). *Classification and regression trees.* Monterey, CA: Wadsworth.

Brennan, R. L. (1983). *Elements of genera lizability.* Iowa City, IA: American College Testing Program.

Brennan, R. L. (2001). *Generalizability theory.* New York: Springer-Verlag.

Brewer, K. (2002). *Combined survey sampling inference: Weighing Basu's elephants.* London: Arnold.

Briggs, C. (1986). *Learning how to ask.* Cambridge, UK: Cambridge University Press.

Broad, W., & Wade, N. (1982). *Betrayers of the truth: Fraud and deceit in the halls of science.* New York: Simon & Schuster.

Brockmeyer, E., Halstrom, H. L., & Jensen, A. (1960). *The life and works of A. K. Erlang.* Kobenhavn: Akademiet for de Tekniske Videnskaber.

Bronfenbrenner, U. (1979). *The ecology of human development.* Cambridge, MA: Harvard University Press.

Bronner, S. E., & Kellner, D. M. (1989). *Critical theory and society: A reader.* New York: Routledge.

Bronshtein, I. N., & Semendyaev, K. A. (1985). *Handbook of mathematics* (K. A. Hirsch, Trans. & Ed.). New York: Van Nostrand Reinhold. (Originally published in 1979)

Brown, C. (1991). *Ballots of tumult: A portrait of volatility in American voting.* Ann Arbor: University of Michigan Press.

Brown, C. (1995). *Chaos and catastrophe theories.* Thousand Oaks, CA: Sage.

Brown, C. (1995). *Serpents in the sand: Essays on the nonlinear nature of politics and human destiny.* Ann Arbor: University of Michigan Press.

Brown, H. I. (1987). *Observation and objectivity.* New York: Oxford University Press.

Brown, S. R., & Melamed, L. E. (1998). *Experimental design and analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-074). Thousand Oaks, CA: Sage.

Brown, S. R., Durning, D. W., & Selden, S. C. (1999). Q methodology. In G. J. Miller & M. L. Whicker (Eds.), *Handbook of research methods in public administration* (pp. 599-637). New York: Dekker.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, *12*, 296-322.

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72-141). Cambridge, UK: Cambridge University Press.

Browne, M. W. (1984). Asymptotic distribution free methods in the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62-83.

Brownlee, K. (1965). *Statistical theory and methodology in science and engineering.* New York: John Wiley.

Bruner, J. (1960). *The process of education.* Cambridge, MA: Harvard University Press.

Bruner, J. (1966). *Toward a theory of instruction.* Cambridge, MA: Harvard University Press.

Brunk, G. G., Caldeira, G. A., & Lewis-Beck, M. S. (1987). Capitalism, socia-lism, and democracy: An empirical inquiry. *European Journal of Political Research*, *15*, 459-470.

Brunswik, E. (1943). Organismic achievement and environmental probability. *The Psychological Review*, *50*, 255-272.

Bryant, C. G. A. (1985). *Positivism in social theory and research.* London: Macmillan.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models.* Newbury Park, CA: Sage.

Bryman, A. (2001). *Social research methods.* Oxford, UK: Oxford University Press.

Bryman, A., & Cramer, D. (1997). *Quantitative data analysis with SPSS for Windows: A guide for social scientists.* New York: Routledge Kegan Paul.

Bryun, T. S. (1966) *The human perspective in sociology: The methodology of participant observation.* Englewood Cliffs, NJ: Prentice Hall.

Bulmer, M. (Ed.). (1982). *Social research ethics.* London: Macmillan.

Bulmer, M., & Solomos, J. (Eds.). (2003). *Researching race and racism.* London: Routledge Kegan Paul.

Bunge, M. (1959). *Causality: The place of the causal principle in modern science.* Cambridge, MA:

Harvard University Press.

Burawoy, M. (1998). The extended case method. *Sociological Theory*, *16*(1), 4-33.

Burgess, R. (1984). *In the field*. London: Allen & Unwin.

Burgess, R. G. (Ed.). (1982). *Field research: A sourcebook and field manual*. London: Allen & Unwin.

Burkhart, R. E. (2000). Economic freedom and democracy: Post-cold war tests. *European Journal of Political Research*, *37*, 237-253.

Burkhart, R. E., and Lewis-Beck, M. S. (1994). Comparative democracy: The economic development thesis. *American Political Science Review*, *88*, 903-910.

Burr, V. (1995). *An introduction to social constructionism*. London: Routledge.

Burton, M. L. (1975). Dissimilarity measures for unconstrained sorting data. *Multivariate Behavioral Research*, *10*, 409-424.

Bury, M. (2001). Illness narratives: Fact or fiction? *Sociology of Health and Illness*, *23*(3), 263-285.

Buttny, R., & Morris, G. H. (2001). Accounting. In W. P. Robinson & H. Giles (Eds.), *The new handbook of language and social psychology* (pp. 285-301). Chichester, England: Wiley.

Byrne, D. (1998). *Complexity theory and the social sciences: An introduction*. London: Routledge.

Byrne, D. (2002). *Interpreting quantitative data*. London: Sage.

Byrne, D. S., McCarthy, P., Harrison, S., & Keithley, J. (1986). *Housing and health*. Aldershot, UK: Gower.

Cacioppo, J. T., Petty, R. E., Losch, M. E., & Kim, H. S. (1986). Electromyographic specificity during simple physical and attitudinal tasks: Location and topographical features of integrated EMG responses. *Biological Psychology*, *18*, 85-121.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2000). *Handbook of psychophysiology* (2nd ed.). Cambridge, UK: Cambridge University Press.

Cahill, S. (1987). Children and civility: Ceremonial deviance and the acquisition of ritual competence. *Social Psychology Quarterly*, *50*, 312-321.

Cain, C. (1991). Personal stories: Identity acquisition and self-understanding in Alcoholics Anonymous. *Ethos*, *19*, 210-253.

Cale, G. (2001). *When resistance becomes reproduction: A critical action research study*. Proceedings of the 42nd Adult Education Research Conference. East Lansing: Michigan State University.

Callendar, J. C., & Osburn, H. G. (1977). A method for maximizing split-half reliability coefficients. *Educational and Psychological Measurement*, *37*, 819-825.

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data*. New York: Cambridge University Press.

Campbell, D. T. (1969). Reforms as experiments. *American Psychologist*, *24*, 409-429.

Campbell, D. T. (1988). *Methodology and epistemology for social science: Selected papers* (E. S. Overman, Ed.). Chicago: University of Chicago Press.

Campbell, D. T. (1991). Methods for the experimenting society. *Evaluation Practice*, *12*, 223-260.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.

Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts*. New York: Guilford.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171-246). Chicago: Rand McNally.

Campbell, D. T., Boruch, R. F., Schwartz, R. D., & Steinberg, J. (1977). Confidentiality-preserving modes of access to files and to interfile exchange for useful statistical analysis. *Evaluation Quarterly*, *1*, 269-300.

Cannell, C., Fowler, J., & Marquis, K. (1968). *The influence of interviewer and respondent psychological and behavioral variables on the reporting of household interviews* (Vital and Health Statistics Series 2, No. 6). Washington, DC: National Center for Health Statistics.

Capdevila, R., & Stainton Rogers, R. (2000). If you go down to the woods today ... : Narratives of Newbury. In H. Addams & J. Proops (Eds.), *Social discourse and environmental policy: An application of Q methodology* (pp. 152-173). Cheltenham, UK: Elgar.

Carless, S. A. (1998). Assessing the discriminant validity of the transformational leadership behaviour as measured by the MLQ. *Journal of Occupational and Organizational Psychology*, *71*, 353-358.

Carley, K., & Prietula, M. (Eds.). (1994). *Computational organization theory*. Hillsdale, NJ: Lawrence Erlbaum.

Carlin, B. P., & Louis, T. A. (1998). *Bayes and empirical Bayes methods for data analysis*. London: Chapman & Hall/CRC.

Carlson, M., & Mulaik, S. A. (1993). Trait ratings from descriptions of behavior as mediated by components of meaning. *Multivariate Behavioral Research*, *28*, 111-159.

Carmines, E. G., & McIver, J. P. (1981). *Unidimensional scaling* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-024). Beverly Hills, CA: Sage.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage.

Carrington, P. J., Scott, J., & Wasserman, S. (Eds.). (2003). *Models and methods in social network analysis*. New York: Cambridge University Press.

Carroll, J. B. (2002). The five-factor personality model: How complete and satisfactory is it? In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 97-126). Mahwah, NJ: Lawrence Erlbaum.

Carroll, J. D., & Arabie, P. (1980). Multidimensional scaling. *Annual Review of Psychology*, *31*, 607-649.

Carroll, J. D., & Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via a N-way generalisation of Eckart-Young decomposition. *Psychometrika*, *35*, 283-299, 310-319.

Carroll, J. M. (1997). Human-computer interaction: Psychology as a science of design. *Annual Review of Psychology*, *48*, 61-83.

Cartwright, N. (1989). *Nature's capacities and their measurement*. Oxford, UK: Oxford University Press.

Casella, G., & Berger, R. L. (1990). *Statistical inference*. Belmont, CT: Duxbury.

Casella, G., & Berger, R. L. (2001). *Statistical inference* (2nd ed.). New York: Wadsworth.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245-276.

Chadhuri, A., & Mukerjee, R. (1987). *Randomized response: Theory and techniques*. New York: Marcel Dekker.

Chalmers, A. (1982). *What is this thing called science?* St. Lucia, Australia: University of Queensland Press.

Chalmers, A. (1999). *What is this thing called science?* (3rd ed.). Buckingham, UK: Open University Press.

Chamberlayne, P., Bornat, J., & Wengraf, T. (2000). *The turn to biographical methods in social science*. London: Routledge Kegan Paul.

Chamberlayne, P., Rustin, M., & Wengraf, T. (Eds.). (2002). *Biography and social exclusion in Europe: Experiences and life journeys*. Bristol, UK: Policy Press.

Charemza, W. W., & Deadman, D. F. (1997). *New directions in econometric practice* (2nd ed.). Cheltenham, UK: Edward Elgar.

Charlton, J., Patrick, D. L., Matthews, G., & West, P. A. (1981). Spending priorities in Kent: A Delphi study. *Journal of Epidemiology and Community Health*, *35*, 288-292.

Charlton, T., Gunter, B., & Hannan, A. (2002). *Broadcast television effects in a remote community*. Hillsdale, NJ: Lawrence Erlbaum.

Charmaz, K. (2000). Grounded theory: Constructivist and objectivist methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 509-535). Thousand Oaks, CA: Sage.

Chatfield, C. (1983). *Statistics for technology* (3rd ed.). London: Chapman & Hall.

Chatfield, C. (1996). *The analysis of time series: An introduction*. New York: Chapman & Hall.

Chavez, L., Hubbell, F. A., McMullin, J. M., Marinez, R. G., & Mishra, S. I. (1995). Structure and meaning in models of breast and cervical cancer risk factors: A comparison of perceptions among Latinas, Anglo women and physicians. *Medical Anthropology Quarterly*, *9*, 40-74.

Checkland, P. B. (1981). *Systems thinking, systems practice.* Chichester, UK: Wiley.

Chell, E. (1998). Critical incident technique. In G. Symon & C. Cassell (Eds.), *Qualitative methods and analysis in organizational research: A practical guide* (pp. 51-72). London: Sage.

Chell, E., & Baines, S. (1998). Does gender affect business performance? A study of micro-businesses in business services in the U. K. *International Journal of Entrepreneurship and Regional Development*, *10*(4), 117-135.

Chen, P. Y., & Popovich, P. M. (2002). *Correlation: Parametric and nonparametric measures.* Thousand Oaks, CA: Sage.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*(2), 367-405.

Cherkassky, V., & Mulier, F. (1998). *Learning from data.* New York: Wiley.

Chernick, M. R. (1999). *Bootstrap methods: A practioner's guide.* New York: Wiley-Interscience.

Cherryholmes, C. H. (1999). *Reading pragmatism.* New York: Teachers College Press.

Chiang, A. C. (1984). *Fundamental methods of mathematical economics* (3rd ed.). New York: McGraw-Hill.

Chiari, G., & Nuzzo, M. L. (1996). Psychological constructivisms: A metatheoretical differentiation. *Journal of Constructivist Psychology*, *9*, 163-184.

Chipman, J. S. (1979). Efficiency of least-squares estimation of linear trend when residuals are autocorrelated. *Econometrica*, *47*, 115-128.

Chouliaraki, L., & Fairclough, N. (1999). *Discourse in late modernity.* Edinburgh, UK: Edinburgh University Press.

Chow, G. (1961). Tests of equality between sets of regression coefficients in linear regression models. *Econometrica*, *28*(3), 591-605.

Chrisman, N. (1997). *Exploring geographic information systems.* New York: Wiley.

Christians, C. (2000). Ethics and politics in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The handbook of qualitative research* (2nd ed., pp. 133-155). Thousand Oaks, CA: Sage.

Cicerelli, V. G., & Associates. (1969). *The impact of Head Start: An evaluation of the effects of Head Start on children's cognitive and affective development* (2 vols.). Athens: Ohio University and Westinghouse Learning Corp.

Cicourel, A. V. (1974). *Theory and method in a critique of Argentine fertility.* New York: John Wiley.

Clark, J. A., & Mishler, E. G. (1992). Attending to patients' stories: Reframing the clinical task. *Sociology of Health and Illness*, *14*, 344-370.

Clarke, H. D., Norpoth, H., & Whiteley, P. F. (1998). It's about time: Modeling political and social dynamics. In E. Scarborough & E. Tanenbaum (Ed.), *Research strategies in the social sciences* (pp. 127-155). Oxford, UK: Oxford University Press.

Clarke, K. A. (2001). Testing nonnested models of international relations: Reevaluating realism. *American Journal of Political Science*, *45*, 724-744.

Clayman, S. E., & Maynard, D. (1995). Ethnomethodology and conversation analysis. In P. ten Have & G. Psathas (Eds.), *Situated order: Studies in the social organization of talk and embodied activities* (pp. 1-30). Washington, DC: University Press of America.

Clayman, S., & Heritage, J. (2002). *The news interview: Journalists and public figures on the air.* Cambridge, UK: Cambridge University Press.

Cleary, T. A., & Linn, R. L. (1969). Error of measurement and the power of a statistical test. *The British Journal of Mathematical and Statistical Psychology*, *22*(1), 49-55.

Clegg, C. W., & Walsh, S. (1998). Soft systems analysis. In G. Symon & C. M. Cassell (Eds.), *Qualitative methods and analysis in organisational research: A practical guide.* London: Sage.

Cleveland, W. S. (1979). Robust locally-weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, *74*, 829-836.

Cleveland, W. S. (1993). *Visualizing data*. Summit, NJ: Hobart.

Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, *83*, 596-610.

Clifford, J., & Marcus, G. E. (Eds.). (1986). *Writing culture: The poetics and politics of ethnography*. Berkeley: University of California Press.

Clogg, C. C. (1978). Adjustment of rates using multiplicative models. *Demography*, *15*, 523-539.

Clogg, C. C. (1982). Some models for the analysis of association in multiway cross-classifications having ordered categories. *Journal of the American Statistical Association*, *77*, 803-815.

Clogg, C. C. (1982). Using association models in sociological research: Some examples. *American Journal of Sociology*, *88*, 114-134.

Clogg, C. C., & Eliason, S. R. (1988). A flexible procedure for adjusting rates and proportions, including statistical methods for group comparisons. *American Sociological Review*, *53*, 267-283.

Clogg, C. C., & Shihadeh, E. S. (1994). *Statistical models for ordinal variables*. Thousand Oaks, CA: Sage.

Clogg, C. C., Shockey, J. W., & Eliason, S. R. (1990). A general statistical framework for adjustment of rates. *Sociological Methods & Research*, *19*, 156-195.

Clough, T. P. (1998). *The end(s) of ethnography: From realism to social criticism* (2nd ed.). New York: Peter Lang.

Cobb, G. W. (1998). *Introduction to design and analysis of experiments*. New York: Springer-Verlag.

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika*, *37*, 256-266.

Cochran, W. G. (1957). Analysis of covariance: Its nature and uses. *Biometrics*, *13*(3), 261-281.

Cochran, W. G. (1965). The planning of observational studies in human populations. *Journal of the Royal Statistical Society, Series A*, *128*, 134-155.

Cochran, W. G. (1953). *Sampling techniques*. New York: Wiley.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York: John Wiley.

Cochrane, D., & Orcutt, G. H. (1949). Application of least squares relationships containing autocorrelated error terms. *Journal of the American Statistical Association*, *44*, 32-61.

Coffey, A., & Atkinson, P. (1996). *Making sense of qualitative data: Complementary research strategies*. Thousand Oaks, CA: Sage.

Cohen, A., Doveh, E., & Eick, U. (2001). Statistical properties of the $r_{WG(J)}$ index of agreement. *Psychological Methods*, *6*, 297-310.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37-46.

Cohen, J. (1978). Partialed products are interactions; partialed powers are curve components. *Psychological Bulletin*, *85*, 858-866.

Cohen, J. (1988). *Statistical power analysis in the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cohen, J. (2001). Smallpox vaccinations: How much protection remains? *Science*, *294*, p. 985.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/ correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Cole, M. (1996). *Cultural psychology: A once and future discipline*. Cambridge, MA: Belknap Press of Harvard University Press.

Coleman, C., & Moynihan, J. (1996). *Understanding crime data: Haunted by the dark figure*. Buckingham, UK: Open University Press.

Collett, D. (1991). *Modelling binary data*. NewYork: Chapman and Hall.

Collican, H. (1999). *Research methods and statistics in psychology*. London: Hodder & Staughton.

Collier, A. (1994). *Critical realism*. London: Verso.

Collins, H. M. (1985). *Changing order: Replication and induction in scientific practice*. London: Sage.

Collins, P. H. (1997). Comment on Heckman's "Truth and method: Feminist standpoint theory revisited": Where's the power? *Signs*, *22*(21),

375-381.

Collins, R. (1995). Prediction in macrosociology: The case of the Soviet collapse. *American Journal of Sociology*, *100*(6), 1552-1593.

Colson, F. (1977). *Sociale indicatoren van enkele aspecten van bevolkingsgroei.* Doctoral dissertation, Katholieke Universiteit Leuven, Department of Sociology, Leuven.

Comstock, D. E. (1994). A method for critical research. In M. Martin & L. C. McIntyre (Eds.), *Readings in the philosophy of social science* (pp. 625-639). Cambridge: MIT Press.

Confidentiality and Data Access Committee. (1999). *Checklist on disclosure potential of proposed data releases.* Washington, DC: Office of Management and Budget, Office of Information and Regulatory Affairs, Statistical Policy Office. Retrieved from www.fcsm.gov/committees/ cdac/checklist_799.doc.

Conger, A. J. (1974). Revised definition for suppressor variables: A guide to their identification and interpretation. *Educational and Psychological Measurement*, *34*, 35-46.

Connell, R. W. (2002). *Gender.* Cambridge, UK: Polity.

Conover, W. J. (1980). *Practical nonparametric statistics.* New York: Wiley.

Conover, W. J. (1998). *Practical nonparametric statistics* (3rd ed.). New York: Wiley.

Conte, R., & Dellarocas, C. (2001). Social order in info societies: An old challenge for innovation. In R. Conte & C. Dellarocas (Eds.), *Social order in multi-agent systems* (pp. 1-16). Boston: Kluwer Academic.

Converse, J. (1987). *Survey research in the United States.* Berkeley: University of California Press.

Converse, J., & Presser, S. (1986). *Survey questions.* Beverly Hills, CA: Sage.

Converse, P. E. (1964). The nature of belief systems in mass publics. In D. Apter (Ed.), *Ideology and discontent* (pp. 206-261). New York: Free Press.

Converse, P. E., & Markus, G. B. (1979). Plus ca change ... : The new CPS Election Study Panel. *American Political Science Review*, *73*, 32-49.

Cook, D. (1996). On the interpretation of regression plots. *Journal of the American Statistical Association*, *91*, 983-992.

Cook, D. R., & Weisberg, S. (1999). *Applied regression including computing and graphics.* New York: Wiley.

Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression.* New York: Chapman and Hall.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand McNally College Publishing.

Cook, T. D., & Payne, M. R. (2002). Objecting to the objections to using random assignment in educational research. In F. Mosteller & R. Boruch (Eds.), *Evidence matters: Randomized trials in education research.* Washington, DC: Brookings Institution.

Cook, T. D., Campbell, D. T., & Peracchio, L. (1990). Quasi-experimentation. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organisational psychology* (2nd ed., Vol. 1, pp. 491-576). Chicago: Rand McNally.

Cooke, B., & Kothari, U. (Eds.). (2001). *Participation: The new tyranny?* London: Zed.

Cooksy, L. J., Gill, P., & Kelly, P. A. (2001). The program logic model as an integrative framework for a multimethod evaluation. *Evaluation and Program Planning*, *24*, 119-128.

Cooley, W., & Lohnes, P. (1971). *Multivariate data analysis.* New York: Wiley.

Coombs, C. H. (1964). *A theory of data.* New York: Wiley.

Cooper, H. (1998). *Synthesizing research: A guide for literature reviews* (3rd ed.). Thousand Oaks, CA: Sage.

Cooper, H., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.

Copi, I. M., & Cohen, C. (1990). *Introduction to logic* (8th ed.). New York: Macmillan.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, *88*(4), 715-730.

Cormack, R. (2001). Population size estimation and

capture-recapture methods. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (Vol. 17). Amsterdam: Elsevier.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2001). *Introduction to algorithms* (2nd ed.). Cambridge: MIT Press.

Correll, S. (1995). The ethnography of an electronic bar: The lesbian cafe. *Journal of Contemporary Ethnography*, *24*(3), 270-298.

Cortazzi, M. (2001). Narrative analysis in ethnography. In P. Atkinson, A. Coffey, S. Delamont, J. Lofland, & L. Lofland (Eds.), *Handbook of ethnography*. London: Sage.

Cortazzi, M., Jin, L., Wall, D., & Cavendish, S. (2001). Sharing learning through narrative communication. *International Journal of Language and Communication Disorders*, *36*, 252-257.

Corter, J. E. (1996). *Tree models of similarity and association* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-112). Thousand Oaks, CA: Sage.

Corti, L. (1993). Using diaries in social research (*Social Research Update*, Iss. 2). Guildford, UK: University of Surrey, Department of Sociology.

Corti, L., Foster, J., & Thompson, P. (1995). *Archiving qualitative research data* (Social Research Update No. 10). Surrey, UK: Department of Sociology, University of Surrey.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*, 98-104.

Costantino, G., Malgady, R. G., & Vazquez, C. (1981). A comparison of the Murray-TAT and a new Thematic Apperception Test for urban Hispanic children. *Hispanic Journal of Behavior Sciences*, *3*, 291-300.

Costigan, P., & Thomson, K. (1992). Issues in the design of CAPI questionnaires for complex surveys. In A. Westlake, R. Banks, C. Payne, & T. Orchard (Eds.), *Survey and statistical computing* (pp. 47-156). London: North Holland.

Costner, H. L. (1965). Criteria for measures of association. *American Sociological Review*, *30*, 341-353.

Couper, M. P. (2000). WEB surveys: A review of issues and approaches. *Public Opinion Quarterly*, *64*, 464-494.

Couper, M. P., Baker, R. P., Bethlehem, J., Clark, C. Z. F., Martin, J., Nichols, W. L., et al. (Eds.). (1998). *Computer assisted survey information collection.* New York: John Wiley.

Courant, Richard. (1988). *Differential and integral calculus* (2 vols.) (E. J. McShane, Trans.). New York: Wiley. (Originally published in 1934)

Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: β is not enough. *Educational and Psychological Measurement*, *61*, 229-248.

Cousins, J. B. (2003). Utilization effects of participatory evaluation. In T. Kellaghan, D. L. Stufflebeam, & L. A. Wingate (Eds.), *International handbook of educational evaluation* (pp. 245-266). Dordrecht, The Netherlands: Kluwer Academic.

Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions in Evaluation*, *80*, 5-23.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society*, *B*, *34*, 187-220.

Cox, D. R., & Oakes, D. (1984). *Analysis of survival data.* New York: Chapman and Hall.

Cox, R. T. (1990). Probability, frequency, and reasonable expectation. In G. Shafer & J. Pearl (Eds.), *Readings in uncertain reasoning* (pp. 353-365). New York: Morgan Kaufmann. (Original work published 1946)

Coxon, A. P. M. (1982). *The user's guide to multidimensional scaling.* London: Heinemann Educational.

Coxon, A. P. M. (1999). *Between the sheets: Sexual diaries and gay men's sex in the era of AIDS.* London: Cassell.

Coxon, A. P. M. (1999). *Sorting data: Collection and analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-127). Thousand Oaks, CA: Sage.

Cramer, J. S. (1986). *Econometric applications of maximum likelihood methods.* Cambridge, UK: Cambridge University Press.

Crano, W. D., & Brewer, M. B. (2002). *Principles*

and methods of social research (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Crapanzano, V. (1980). *Tuhami: Portrait of a Moroccan.* Chicago: University of Chicago Press.

Crenshaw, K., Gotanda, N., Peller, G., & Thomas, K. (Eds.). (1995). *Critical race theory.* New York: The New York Press.

Cressey, D. (1950). The criminal violation of financial trust. *American Sociological Review*, *15*, 738-743.

Cressey, D. (1953). *Other people's money.* Glencoe, IL: Free Press.

Cressie, N. (1993). *Statistics for spatial data.* New York: Wiley.

Cressie, N., & Read, T. (1984). Multinomial goodness of tests. *Journal of Royal Statistical Society Series B*, *46*, 440-464.

Creswell, J. W. (1995). *Research design: Quantitative and qualitative approaches.* Thousand Oaks, CA: Sage.

Crocker, L., & Algina, J. (1986). *Introduction to classical & modern test theory.* New York: Harcourt Brace Jovanovich.

Croll, P. (1986). *Systematic classroom observation.* Lewes, UK: Falmer.

Cromwell, J. B., Labys, W., & Terraza, M. (1994). *Univariate tests for time series models.* Thousand Oaks, CA: Sage.

Cronbach, L. J. (1943). On estimates of test reliability. *Journal of Educational Psychology*, *34*, 485-494.

Cronbach, L. J. (1946). A case study of the split-half reliability coefficient. *Journal of Educational Psychology*, *37*, 473-480.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334.

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin*, *52*, 177-193.

Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: HarperCollins.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability of scores and profiles.* New York: Wiley.

Croon, M. A., Bergsma, W., & Hagenaars, J. A. (2000). Analyzing change in categorical variables by generalized log-linear models. *Sociological Methods & Research*, *29*, 195-229.

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process.* London: Sage.

Crowne, D., & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence.* New York: Wiley.

Cruse, D. A. (1986). *Lexical semantics.* Cambridge, UK: Cambridge University Press.

Cullen, C. G. (1990). *Matrices and linear transformations* (2nd ed.). New York: Dover.

Curtin, R., Presser, S., & Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment. *Public Opinion Quarterly*, *64*, 413-428.

Czaja, R., Blair, J., & Sebestile, J. P. (1982). Respondent selection in a telephone survey: Comparison of three techniques. *Journal of Marketing Research*, *21*, 381-385.

Czyzewski, M. (1994). Reflexivity of actors versus reflexivity of accounts. *Theory, Culture and Society*, *11*, 161-168.

D'Andrade, R. (1995). *The development of cognitive anthropology.* Cambridge, UK: Cambridge University Press.

Dale, A., & Marsh, C. (Eds.). (1993). *The 1991 census user's guide.* London: HMSO.

Dale, A., Arber, S., & Procter, P. (1988). *Doing secondary analysis.* London: Unwin Hyman.

Dale, A., Fieldhouse, E., & Holdsworth, C. (2000). *Analyzing census microdata.* London: Arnold.

Dandridge, T. C., Mitroff, I., & Joyce, W. F. (1980). Organizational symbolism: A topic to expand organizational analysis. *Academy of Management Review*, *5*, 77-82.

Daniel, W. W. (1993). *Collecting sensitive data by randomized response: An annotated bibliography* (2nd ed.). Atlanta: Georgia State University Business Press.

Daniels, H. E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika*, *35*, 129-135.

Darlington, R. B. (1990). *Regression and linear models.* New York: McGraw-Hill.

Darnell, A. C. (1995). *A dictionary of econometrics.* Cheltenham, UK: Edward Elgar.

Darroch, J. N., Lauritzen, S. L., & Speed, T. P. (1980). Markov fields and log-linear models. *Annals of Mathematical Statistics*, *43*, 1470-1480.

David, H. A., & Moeschberg, M. L. (1978). *The theory of competing risks* (Griffin's Statistical Monograph #39). New York: Macmillan.

David, P. (1985). Clio and the economics of QWERTY. *American Economic Review*, *75*, 332-337.

Davidson, J., Hendry, D. F., Srba, F., & Yeo, S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal*, *88*, 661-692.

Davidson, R., & MacKinnon J. G. (1993). *Estimation and inference in econometrics.* New York: Oxford University Press.

Davis, J. A. (1985). *The logic of causal order* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-055). Beverly Hills, CA: Sage.

Davis, J. A., & Smith, T. W. (1992). *The NORC General Social Survey: A user's guide.* Newbury Park, CA: Sage.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application.* Cambridge, UK: Cambridge University Press.

Day, N. E. (1969). Estimating the components of a mixture of two normal distributions. *Biometrika*, *56*, 463-474.

De Beauvoir, S. (1970). *The second sex.* New York: Alfred A. Knopf. (Original work published 1949)

de Leeuw, J., & Kreft, I. G. G. (Eds.). (in press). *Handbook of quantitative multilevel analysis.* Dordrecht, The Netherlands: Kluwer Academic.

de Leeuw, J., & van der Heijden, P. G. M. (1988). The analysis of time-budgets with a latent time-budget model. In E. Diday (Ed.), *Data analysis and informatics* (Vol. 5, pp. 159-166). Amsterdam: North-Holland.

de Saussure, F. (1979). *Cours de linguistique générale* (T. de Mauro, Ed.). Paris: Payot.

de Vaus, D. (Ed.). (2002). *Social surveys* (4 vols.). London: Sage.

de Vaus, D. A. (1995). *Surveys in social research* (4th ed.). Sydney, Australia: Allen & Unwin.

de Vaus, D. A. (2001). *Research design in social research.* London: Sage.

Deacon, D., Bryman, A., & Fenton, N. (1998). Collision or collusion? A discussion of the unplanned triangulation of quantitative and qualitative research methods. *International Journal of Social Research Methodology*, *1*, 47-63.

Deacon, D., Pickering, M., Golding, P., & Murdock, G. (1999). *Researching communications.* London: Arnold.

DeBoef, S., & Granato, J. (2000). Testing for cointegrating relationships with near-integrated data. *Political Analysis*, *8*, 99-117.

Deckner, D. F., Adamson, L. B., & Bakeman, R. (2003). Rhythm in mother-infant interactions. *Infancy*, *4*, 201-217.

DeGroot M. H., & Schervish, M. J. (2002). *Probability and statistics* (3rd ed.). Reading, MA: Addison-Wesley.

Dehejia, R. H., & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, *94*, 1053-1062.

Del Monte, K. (2000). Partners in inquiry: Ethical challenges in team research. *International Social Science Review*, *75*, 3-14.

Delanty, G. (1997). *Social science: Beyond constructivism and realism.* Buckingham, UK: Open University Press.

DeMaris, A. (1992). *Logit modeling.* Newbury Park, CA: Sage.

Deming, W. E. (1997). *Statistical adjustment of data.* New York: Dover.

DeNavas-Walt, C., & Cleveland, R. W. (2002). *Money income in the United States* (United States Census Bureau, Current Population Reports, P60-218). Washington, DC: Government Printing Office. Available: www.census.gov/prod/2002pubs/p60-218.pdf.

Denzin, N. (1983). Interpretive interactionism. In G. Morgan (Ed.), *Beyond method: Strategies for social research* (pp. 128-142). Beverly Hills, CA: Sage.

Denzin, N. (1992). *Symbolic interactionism: The politics of interpretation.* Oxford, UK: Blackwell.

Denzin, N. K. (1970). *The research act in sociology.* Chicago: Aldine.

Denzin, N. K. (1989). *Interpretive biography.* Newbury Park, CA: Sage.

Dezin, N. K., (1989). *Interpretive interactionism.* Newbury Park, CA: Sage.

Denzin, N. K., (1997). *Interpretive enthnography.* Thousand Oaks, CA: Sage.

Denzin, N. K., & Lincoln, Y. S. (2000). The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 1-28). Thousand Oaks, CA: Sage.

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2000). *Handbook of qualitative research* (2nd ed.). Thousand Oaks, CA: Sage.

Denzin, N. K. & Lincoln, Y. S. (2000). Introduction: The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 1-28). Thousand Oaks, CA: Sage.

Derrida, J. (1976). *Of grammatology* (G. Spivak, Trans.). Baltimore, MD: Johns Hopkins University Press.

Derrida, J. (1995). *The gift of death.* Chicago: The University of Chicago Press.

Derrida, J., & Ferraris, F. (2001). *A taste for the secret.* Cambridge, UK: Polity.

DeVellis, R. F. (1991). *Scale development: Theory and applications.* Newbury Park, CA: Sage.

Devore, J. L. (1991). *Probability and statistics for engineering and the sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Dexter, L. A. (1970). *Elite and specialized interviewing.* Evanston, IL: Northwest University Press.

Dey, I. (1993). *Qualitative data analysis.* London: Routledge Kegan Paul.

Diamond, S. S. (2000). Reference guide on survey research, in *Reference manual on scientific evidence.* (2nd ed., pp. 229-276). Washington, D. C.: Federal Judicial Center.

Dickens, P. (2000). *Social Darwinism.* Buckingham, UK: Open University Press.

Diebold, F. X. (2001). *Elements of forecasting* (2nd ed.). Cincinnati, OH: South-Western.

Dienstbier, R. A. (1989). Arousal and physiological toughness: Implications for mental and physical health. *Psychological Review*, *96*, 84-100.

Diesing, P. (1991). *How does social science work?* Pittsburgh, PA: University of Pittsburgh Press.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, *41*, 417-440.

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method.* New York: Wiley.

Dillman, D. A. (2000). *Mail and Internet surveys: The tailored design method.* New York: Wiley.

Dilthey, W. (1985). *Poetry and experience: Selected works* (Vol. 5). Princeton, NJ: Princeton University Press.

Dion, K. K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*, 285-290.

DiPrete, T. A., & Forristal, J. D. (1994). Multilevel models: Methods and substance. *Annual Review of Sociology*, *20*, 331-357.

Dixon-Woods, M., Fitzpatrick, R., & Roberts, K. (2001). Including qualitative research in systematic reviews: Problems and opportunities. *Journal of Evaluation in Clinical Practice*, *7*, 125-133.

Doksum, K. A., & Sievers, G. L. (1976). Plotting with confidence: Graphical comparisons of two populations. *Biometrika*, *63*, 421-434.

Doreian, P., Batagelj, V., & Ferligoj, A. (1994). Partitioning networks based on generalized concepts of equivalence. *Journal of Mathematical Sociology*, *19*, 1-27.

Douglas, J. D. (1985). *Creative interviewing.* Beverly Hills, CA: Sage.

Douglass, B., & Moustakas, C. (1985). Heuristic inquiry: The internal search to know. *Journal of Humanistic Psychology*, *25*, 39-55.

Dovidio, J. F., Kawakami, K., & Beach, K. R. (2001). Implicit and explicit attitudes:

Examination of the relationship between measures of intergroup bias. In R. Brown & S. L. Gaertner (Eds.), *Blackwell handbook of social psychology: Intergroup processes* (pp. 175-197). Malden, MA: Blackwell.

Downs, A. (1957). *An economic theory of democracy.* New York: HarperCollins.

Doyle, P., Martin, B., & Moore, J. (2000). *Improving income measurement.* The Survey of Income Program Participation (SIPP) Methods Panel. Washington, DC: U. S. Bureau of the Census.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). New York: Wiley.

Dreher, A. U. (2000). *Foundations for conceptual research in psychoanalysis* (Psychoanalytic Monograph 4). London: Karnac.

Drew, P., & Heritage, J. (Eds.). (1992). *Talk at work: Interaction in institutional settings.* Cambridge, UK: Cambridge University Press.

Dryzek, J. S., & Holmes, L. T. (2002). *Post-communist democratization.* Cambridge, UK: Cambridge University Press.

Dubin, J. A., & Rivers, D. (1989). Selection bias in linear regression, logit and probit models. *Sociological Methods & Research*, *18*, 360-390.

Duda, R., Hart, P. E., & Stork, D. G. (2001). *Pattern classification.* New York: Wiley.

Dukes, R. L., Ullman, J. B., & Stein, J. A. (1995). An evaluation of D.A.R.E. (Drug Abuse Resistance Education), using a Solomon four-group design with latent variables. *Evaluation Review*, *19*(4), 409-435.

Duncan, O. D. (1966). Path analysis: Sociological examples. *American Journal of Sociology*, *72*, 1-16.

Duncan, O. D. (1975). *Introduction to structural equation models.* New York: Academic Press.

Duncan, O. D., Haller, A., & Portes, A. (1968). Peer influence on aspiration: A reinterpretation. *American Journal of Sociology*, *75*, 119-137.

Dupré, J. (2001). *Human nature and the limits of science.* Oxford, UK: Oxford University Press.

Dupré, J., & Cartwright, N. (1988). Probability and causality: Why Hume and indeterminism don't mix. *Nous*, *22*, 521-536.

Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regressions I. *Biometrika*, *37*, 409-428.

Durkheim, E. (1951). *Suicide.* Glencoe, IL: Free Press.

Durkheim, E. (1952). *Suicide.* London: Routledge Kegan Paul. (Original work published 1896)

Durkheim, E. (1964). *The rules of scientific method.* Glencoe, IL: Free Press.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 276-284.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good but ... A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*, 109-128.

Easterby-Smith, M., Thorpe, R., & Holman, D. (1996). Using repertory grids in management. *Journal of European Industrial Training*, *20*, 1-30.

Easterby-Smith, M., Thorpe, R., & Lowe, A. (2002). *Management research: An introduction* (2nd ed.). London: Sage.

Eatwell, J., Milgate, M., & Newman, P. (1987). *The new Palgrave: A dictionary of economics.* London: Macmillan.

Eckstein, H. (1975). Case study and theory in political science. In F. I. Greenstein & N. W. Polsby (Eds.), *Handbook of political science: Vol. 7, Strategies of inquiry* (pp. 79-137). Reading, MA: Addison-Wesley.

Eco, U. (1972). Towards a semiotic inquiry into the television message. In *Working papers in cultural studies* (Vol. 3, pp. 103-121). Birmingham: Centre for Contemporary Cultural Studies. (Original work published 1965)

Eco, U. (1976). *A theory of semiotics.* Bloomington: Indiana University Press.

Eco, U. (1981). *The role of the reader: Explorations in the semiotics of texts.* London: Hutchinson.

Economic and Social Research Council Research Centre on Micro-Social Change. (2001, February 28). *British Household Panel Survey* [computer file] (Study Number 4340). Colchester, UK: The Data Archive [distributor].

Eden, D. (1990). *Pygmalion in management:*

Productivity as a self-fulfilling prophecy. Lexington, MA: D. C. Heath.

Edwards, A. (1948). On Guttman's scale analysis. *Educational and Psychological Measurement*, *8*, 313-318.

Edwards, A. L. (1957). *Techniques of attitude construction*. New York: Appleton- Century-Crofts.

Edwards, A. W. F. (1972). *Likelihood: An account of the statistical concept of likelihood and its application to scientific inference*. Cambridge, UK: Cambridge University Press.

Edwards, D. (1997). *Discourse and cognition*. London: Sage.

Edwards, D. (2000). *Introduction to graphical modelling* (2nd ed.). New York: Springer-Verlag.

Edwards, D., & Potter, J. (1992). *Discursive psychology*. London: Sage.

Edwards, D., Ashmore, M., & Potter, J. (1995). Death and furniture: The rhetoric, politics, and theology of bottom line arguments against relativism. *History of the Human Sciences*, *8*, 25-49.

Edwards, W. S., Winn, D. M., Kurlantzick, V. et al. (1994). *Evaluation of National Health Interview Survey diagnostic reporting*. National Center for Health Statistics. Vital Health Stat 2(120).

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, *7*, 1-26.

Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans*. Philadelphia: Society for Industrial and Applied Mathematics.

Efron, B. (1986). Why isn't everyone a Bayesian? *The American Statistician*, *40*, 1-5.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Egger, M., Davey Smith, G., & Altman, D. G. (2001). *Systematic reviews in health care: Meta-analysis in context* (2nd ed.). London: BMJ Books.

Egger, M., Smith, G., Schneider, M., & Bias in meta-analysis detected by a simple, graphical test. Minder, C. (1997). *British Medical Journal*, *315*, 629-634.

Eibl-Eibesfeldt, I. (1989). *Human ethology*. Hawthorne, NY: Aldine de Gruyter.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*, 532-550.

Eisenhart, M. (1998). On the subject of interpretive reviews. *Review of Educational Research*, *68*(4), 391-399.

Eisner, E. W. (1997). The new frontier in qualitative research methodology. *Qualitative Inquiry*, *3*, 259-273.

Elder, G., Pavalko, E. K., & Clipp, E. C. (1993). *Working with archival lives: Studying lives* (Sage University Papers on Quantitative Applications on the Social Sciences, 07-088). Newbury Park, CA: Sage.

Eliason, S. R. (1993). *Maximum likelihood estimation: Logic and practice*. Newbury Park, CA: Sage.

Ellis, B. D., & Lierse, C. (1994). Dispositional essentialism. *Australasian Journal of Philosophy*, *72*(1), 27-45.

Ellis, C. (1995). *Final negotiations: A story of love, loss, and chronic illness*. Philadelphia: Temple University Press.

Ellis, C. (2003). *The ethnographic "I": A methodological novel on doing autoethnography*. Walnut Creek, CA: AltaMira.

Ellis, C. S. (1991). Sociological introspection and emotional experience. *Symbolic Interaction*, *14*, 23-50.

Ellis, C., & Bochner, A. P. (2000). Autoethnography, personal narrative, reflexivity: Researcher as subject. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 733-768). Thousand Oaks, CA: Sage.

Ellis, C., & Bochner, A. P. (Eds.). (1996). *Composing ethnography: Alternative forms of qualitative writing*. Walnut Creek, CA: AltaMira.

Elmasri, R. A., & Navathe, S. B. (2001). *Fundamentals of database systems*. New York: Addison-Wesley.

Elster, J. (1989). *Nuts and bolts for the social sciences*. Cambridge, UK: Cambridge University Press.

Elster, J. (1989). *The cement of society: A study of social order*. Cambridge, UK: Cambridge University Press.

Elster, J. (Ed.). (1986). *Rational choice*. Oxford,

UK: Basil Blackwell.

Ember, C. R., & Ember, M. (2001). *Cross-cultural research methods.* Walnut Creek, CA: AltaMira.

Embree, L., Behnke, E. A., Carr, D., Evans, J. C., Huertas-Jourda, J., Kockelmans, J. J., et al. (Eds.). (1996). *The encyclopedia of phenomenology.* Dordrecht, The Netherlands: Kluwer.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Lawrence Erlbaum.

Emerson, R. M., & Pollner, M. (1988). On the use of members' responses to researchers' accounts. *Human Organization*, *47*, 189-198.

Emerson, R. M., Fretz, R. I., & Shaw, L. L. (1995). *Writing ethnographic fieldnotes.* Chicago: University of Chicago Press.

Emirbayer, M., & Mische, A. (1998). What is agency? *American Journal of Sociology*, *103*(4), 962-1023.

Emmison, M., & Smith, P. (2000). *Researching the visual.* London: Sage.

Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica*, *55*, 251-276.

Engle, R. F., Granger, C. W. J. (1991). *Long run relationships: Readings in cointegration.* New York: Oxford University Press.

Engle, R. F., Hendry, D. F., & Richard, J. F. (1983). Exogeneity. *Econometrica*, *51*, 277-304.

Epley, N., & Huff, C. (1998). Suspicion, affective response, and educational benefit of deception in psychology research. *Personality and Social Psychology Bulletin*, *24*, 759-768.

Epstein, J. M., & Axtell, R. (1996). *Growing artificial societies: Social science from the bottom up.* Cambridge: MIT Press.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge: MIT Press.

Erikson, E. H. (1975). *Life history and the historical moment.* New York: Norton.

Erikson, R., Goldthorpe, J. H., & Portocarero, L. (1979). Intergenerational mobility in three Western European societies. *British Journal of Sociology*, *30*.

Erlang, A. K. (1935). *Fircifrede logaritmetavler og andre regnetavler til brug ved undervisning og i praksis.* Kobenhavn: G.E.C. Gads.

Ermarth, M. (1978). *Wilhelm Dilthey: The critique of historical reason.* Chicago: University of Chicago Press.

Escofier, B., & Pagès, J. (1988). *Analyses factorielles multiples* [Multiple factor analyses]. Paris: Dunod.

Essed, Ph., & Goldberg, T. D. (Eds.). (2002). *Race critical theories: Text and context.* Malden, MA: Blackwell.

Eubank, R. L. (1988). Quantiles. In S. Kotz, N. L. Johnson, & C. B. Read (Eds.), *Encyclopedia of statistical sciences* (Vol. 7, pp. 424-432). New York: Wiley.

European Social Survey Central Co-ordinating Team. (2001). *European Social Survey (ESS): Specification for participating countries.* London: Author.

Eurostat. (1998). *Labour force survey: Methods and definitions* (1998 ed.). Luxembourg: Author.

Evans, J. L. (2000). *Early childhood counts.* Washington, DC: World Bank.

Evans, M., Hastings, N., & Peacock, B. (2000). *Statistical distributions* (3rd ed.). New York: Wiley.

Everitt, B. S. (1992). *The analysis of contingency tables.* London: Chapman & Hall.

Everitt, B., Landau, S., & Leese, M. (2001). *Cluster analysis* (4th ed.). Oxford, UK: Oxford University Press.

Exner, J. E. (2002). *The Rorschach: A comprehensive system* (Vol. 1). New York: John Wiley.

Fairclough, N. (2000). *New labour, new language?* London: Routledge Kegan Paul.

Fairclough, N., & Wodak, R. (1997). Critical discourse analysis. In T. van Dijk (Ed.), *Discourse as social interaction.* Thousand Oaks, CA: Sage.

Fals-Borda, O., & Anisur Rahman, M. (Eds.). (1991). *Action and knowledge: Breaking the monopoly with participatory action-research.* New York: Apex.

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications.* London: Chapman & Hall.

Faugier, J., & Sargeant, M. (1997). Sampling hard to reach populations. *Journal of Advanced Nursing*, *26*, 790-797.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, *54*, 297-327.

Featherman, D. L., & Hauser, R. M. (1978). *Opportunity and change.* New York: Academic Press.

Featherstone, M. (1988). In pursuit of the postmodernism: An introduction. *Theory, Culture & Society*, *5*, 195-215.

Fechner, G. T. (1860). *Elemente der Psychophysik.* Leipzig: Breitkopf & Härtel.

Fechner, G. T. (1877). *In sachen der Psychophysik.* Leipzig: Breitkopf & Härtel.

Federal Committee on Statistical Methodology. (1994). *Report on statistical disclosure limitation methodology* (Statistical Policy Working Paper # 22). Prepared by the Subcommittee on Disclosure Limitation Methodology. Washington, DC: Office of Management and Budget, Office of Information and Regulatory Affairs, Statistical Policy Office. Retrieved from www. fcsm. gov/working-papers/wp22.html

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 105-146). New York: Macmillan.

Feller, W. (1968). *An introduction to probability theory and its applications* (3rd ed.). New York: John Wiley.

Ferrell, J., & Hamm, M. S. (1998). True confessions: Crime, deviance, and field research. In J. Ferrell & M. Hamm (Eds.), *Ethnography at the edge: Crime, deviance and field research* (pp. 2-19). Boston: Northeastern University Press.

Fetterman, D. M. (1998). *Ethnography: Step by step* (2nd ed.). Thousand Oaks, CA: Sage.

Fetterman, D. M. (2002). Web surveys to digital movies: Technological tools of the trade. *Educational Researcher*, *31*(6), 29-37.

Feyerabend, P. (1978). *Against method.* London: Verso.

Feynman, R. P., & Weinberg, S. (1987). *Elementary particles and the laws of physics: The 1986 Dirac Memorial Lectures.* Cambridge, UK: Cambridge University Press.

Field, A. (1998). A bluffer's guide to ... sphericity. *British Psychological Society: Mathermatical, Statistical & Computing Newsletter*, *6*, 13-22.

Fielding, J. (1993). Coding and managing data. In N. Gilbert (Ed.), *Researching social life* (pp. 218-238). Thousand Oaks, CA: Sage.

Fielding, N. G., & Lee, R. M. (1998). *Computer analysis and qualitative research.* London: Sage.

Fienberg, S. E. (1977). *The analysis of cross-classified categorical data.* Cambridge: MIT Press.

Fienberg, S. E. (1980). *The analysis of cross-classified categorical data* (2nd ed.). Cambridge: MIT Press.

Fillmore, C. (1975). An alternative to checklist theories of meaning. In C. Cogen, H. Thompson, G. Thurgood, K. Whistler, & J. Wright (Eds.), *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society* (pp. 123-131). Berkeley, CA: Berkeley Linguistics Society.

Filmer, P. (1998). Analysing literary texts. In C. Seale (Ed.), *Researching society and culture.* London: Sage.

Fink, A., & Kosecoff, J. (1988). *How to conduct surveys: A step-by-step guide.* Newbury Park, CA: Sage.

Firebaugh, G. (2001). Ecological fallacy. *International encyclopedia for the social and behavioral sciences* (Vol. 6, pp. 4023-4026). Oxford, UK: Pergamon.

Fisher, C. B., & Wallace, S. A. (2000). Through the community looking glass: Reevaluating the ethical and policy implications of research on adolescent risk and psychopathology. *Ethics & Behavior*, *10*(2), 99-118.

Fisher, R. A. (1925). *Statistical methods for research workers* (1st ed.). Edinburgh, UK: Oliver & Boyd.

Fisher, R. A. (1925). Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society*, *22*, 700-725.

Fisher, R. A. (1934). Two new properties of mathematical likelihood. *Proceedings of the Royal Society of London*, *Series A*, *144*, 285-307.

Fisher, R. A. (1935). *The design of experiments* (1st ed.). Edinburgh, UK: Oliver and Boyd.

Fisher, R. A. (1971). *Design of experiments.* New York: Hafner Press. (Original work published 1935).

Fisher, R. A., & Yates, F. (1934). The six by six Latin squares. *Proceedings of the Cambridge Philosophical Society*, *30*, 492-507.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, *51* (4), 327-358.

Flick, U. (2002). *An introduction to qualitative research* (2nd ed.). London: Sage.

Foley, D. E. (1990). *Learning capitalist culture: Deep in the heart of Texas.* Philadelphia: University of Pennsylvania Press.

Fontana, A., & Frey, J. H. (1994). Interviewing: The art of science. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 361-376). Thousand Oaks, CA: Sage.

Fontana, A., & Frey, J. H. (2000). The interview: From structured questions to negotiated text. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 645-672). Thousand Oaks, CA: Sage.

Forrester, J. W. (1961). *Industrial dynamics.* Cambridge: MIT Press.

Forster, E., & McCleery, A. (1999). Computer assisted personal interviewing: A method of capturing sensitive information. *IASSIST Quarterly*, *23*(2), 26-38.

Foster, J., & Sheppard, J. (1995). *British archives: A guide to archival resources in the United Kingdom* (3rd ed.). London: Macmillan.

Foucault, M. (1985). *Power/knowledge: Selected interviews and other writings: 1972-1977* (C. Gordon, Ed.). New York: Pantheon.

Foucault, M. (1991). Questions of method. In G. Burchell, C. Gordon, & P. Miller (Eds.), *The Foucault effect: Studies in governmentality* (pp. 73-86). London: Harvester Wheatsheaf.

Foucault, M. (1991). *Remarks on Marx: Conversations with Duccio Trombadori* (R. J. Goldstein & J. Cascaito, Trans.). New York: Semiotext(e).

Foucault, M. (1995). *Discipline and punish: The birth of the prison.* New York: Vintage Books.

Foucault, M. (1998). *The will to knowledge: The history of sexuality* (Vol. 1). London: Penguin.

Foucault, M. (2002). *Archeology of knowledge.* London: Routledge.

Fowler, F. J., Jr. (1988). *Survey research methods* (Rev. ed.). Newbury Park, CA: Sage.

Fowler, F. J., Jr. (1995). *Improving survey questions: Design and evaluation.* Thousand Oaks, CA: Sage.

Fowler, F. J., Jr. (2002). *Survey research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Fowler, F. J., Jr., & Mangione, T. W. (1990). *Standardized survey interviewing.* Newbury Park, CA: Sage.

Fowler, R. L. (1985). Point estimates and confidence intervals in measures of association. *Psychological Bulletin*, *98*, 160-165.

Fowler, R. L. (1985). Testing for substantive significance in applied research by specifying non-zero effect nullhypotheses. *Journal of Applied Psychology*, *70*, 215-218.

Fox, J. (1991). *Regression diagnostics* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-079). Newbury Park, CA: Sage.

Fox, J. (1997). *Applied regression analysis, linear models, and related methods.* Thousand Oaks, CA: Sage.

Fox, J. (2000). *Nonparametric simple regression: Smoothing scatterplots.* Thousand Oaks, CA: Sage.

Fox, J. (2000). *Multiple and generalized nonparametric regression.* Thousand Oaks, CA: Sage.

Fox, J. (2002). *An R and S-PLUS companion to applied regression.* Thousand Oaks, CA: Sage.

Fox, J. A., & Tracy, P. E. (1986). *Randomized response: A method for sensitive surveys.* Beverly Hills, CA: Sage.

Fox, J., & Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association*, *87*, 178-183.

Fox, W. (1998). *Social statistics* (3rd ed.). Bellevue, WA: MicroCase.

Frank, I. E., & Friedman, J. H. (1993). A statistical view of chemometrics regression tools. *Technometrics*, *35*, 109-148.

Frank, O., & Strauss, D. (1986). Markov graphs.

Journal of the American Statistical Association, *81*, 832-842.

Frankfort-Nachmias, C., & Nachmias, D. (1996). *Research methods in the social sciences* (5th ed.). New York: St. Martin's.

Frankfort-Nachmias, C., & Nachmias, D. (2000). *Research methods in the social sciences* (6th ed.). New York: Worth.

Fransella, F., & Bannister, D. (1977). *A manual for repertory grid technique*. London: Academic Press.

Franses, P. H. (1998). *Time series models for business and economic forecasting*. Cambridge, UK: Cambridge University Press.

Franzese, R. (2002). *Macroeconomic policy of developed democracies*. Cambridge, UK: Cambridge University Press.

Franzosi, R. (2003). *From words to numbers*. Cambridge, UK: Cambridge University Press.

Fraser, N. (1997). *Justice interruptus: Critical reflections on the "postsocialist" condition*. London: Routledge Kegan Paul.

Frederick, B. N. (1999). Partitioning variance in the multivariate case: A step-by-step guide to canonical commonality analysis. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 5, pp. 305-318). Stamford, CT: JAI.

Frederick, R. I., & Crosby, R. D. (2000). Development and validation of the Validity Indicator Profile. *Law and Human Behavior*, *24*, 59-82.

Freedman, D. A. (2001). Ecological inference and the ecological fallacy. *International encyclopedia for the social and behavioral sciences* (Vol. 6, pp. 4027-4030). Oxford, UK: Pergamon.

Freedman, D. A., & Wachter, K. W. (2001). *On the likelihood of improving the accuracy of the census through statistical adjustment* (Tech. Rep. 612). Berkeley: University of California, Department of Statistics.

Freedman, D. A., Klein, S. P., Ostland, M., & Roberts, M. R. (1998). Review of "A solution to the ecological inference problem." *Journal of the American Statistical Association*, *93*, 1518-1522. (Discussion appears in Vol. 94, pp. 352-357)

Freedman, D. A., Pisani, R., & Purves, R. A. (1998). *Statistics*. 3rd ed. New York: W. W. Norton, Inc.

Freeman, D. (1983). *Margaret Mead and Samoa*. Cambridge, MA: Harvard University Press.

Freeman, J. F. (1983). Granger causality and the time series analysis of political relationships. *American Journal of Political Science*, *27*, 325-355.

Frey, J. H. (1989). *Survey research by telephone* (2nd ed.). Newbury Park, CA: Sage.

Frey, J. H., & Fontana, A. (1991). The group interview in social research. *The Social Science Journal*, *28*, 175-187.

Frey, J. H., & Oishi, S. M. (1995). *How to conduct interviews by telephone and in person*. Thousand Oaks, CA: Sage.

Fricker, R. D., & Schonlau, M. (2002). Advantages and disadvantages of Internet research surveys: Evidence from the literature. *Field Methods*, *14* (4), 347-365.

Friedkin, N. (1998). *A structural theory of social influence*. New York: Cambridge University Press.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, *32*, 675-701.

Friedman, M. (1962). The interpolation of time series by related series. *Journal of the American Statistical Association*, *57*, 729-757.

Friedman, M. (1991). The re-evaluation of logical positivism. *Journal of Philosophy*, *88* (10), 505-519.

Friedricks, R. (1970). *A sociology of sociology*. New York: Free Press.

Friendly, M. (1992). Mosaic displays for loglinear models. In American Statistical Association (Ed.), *Proceedings of the Statistical Graphics Section* (pp. 61-68). Alexandria, VA: American Statistical Association.

Friendly, M. (1994). Mosaic displays for multi-way contingency tables. *Journal of the American Statistical Association*, *89*, 190-200.

Friendly, M. (1998). Mosaic displays. In S. Kotz, C. Reed, & D. L. Banks (Eds.), *Encyclopedia of statistical sciences* (Vol. 2, pp. 411-416). New York: John Wiley.

Friendly, M. (1999). Extending mosaic displays: Marginal, conditional, and partial views of categorical data. *Journal of Computational and*

Graphical Statistics, *8*(3), 373-395.

Friendly, M. (2000). *Visualizing categorical data.* Cary, NC: SAS Institute.

Friendly, M. (2002). A brief history of the mosaic display. *Journal of Computational and Graphical Statistics*, *11*(1), 89-107.

Fu, V., Mare, R., & Winship, C. (2003). Sample selection bias models. In M. A. Hardy & A. Bryman (Eds.), *Handbook of data analysis.* London: Sage.

Fuller, D. (1999). Part of the action, or "going native"? Learning to cope with the politics of integration. *Area*, *31*(3), 221-227.

Fuller, W. A. (1987). *Measurement error models.* New York: John Wiley.

Fuss, D. (1989). *Essentially speaking.* New York: Routledge. Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, *58*, 453-467.

Gadamer, H.-G. (1975). *Truth and method.* New York: Seabury.

Gadamer, H-G. (1989). *Truth and method* (rev. 2nd ed.). New York: Crossroad.

Gaito, J. (1980). Measurement scales and statistics: Resurgence of an old misconception. *Psychological Bulletin*, *87*, 564-567.

Gallmeier, C. P. (1991). Leaving, revisiting, and staying in touch: Neglected issues in field research. In W. B. Shaffir & R. A. Stebbins (Eds.), *Fieldwork experience: Qualitative approaches to social research* (pp. 224-231). London: Sage.

Galton, F. (1886). Regression toward mediocrity in hereditary stature. *Journal of the Anthropological Institute*, *15*, 246-263.

Galton, F. (1889) *Natural inheritance.* London: Macmillan.

Galton, M., Simon, B., & Croll, P. (1980). *Inside the primary classroom.* London: Routledge and Kegan Paul.

Galunic, D. C., & Eisenhardt, K. M. (2001). Architectural innovation and modular corporate forms. *Academy of Management Journal*, *44*(6), 1229-1249.

Game, A. (1991). *Undoing the social: Towards a deconstructive sociology.* Buckingham, UK: Open University Press.

Garfinkel, H. (1967). *Studies in ethnomethodology.* Englewood Cliffs, NJ: Prentice Hall.

Garfinkel, H. (1974). On the origins of the term "ethnomethodology." In R. Turner (Ed.), *Ethnomethodology* (pp. 15-18). Harmondsworth, UK: Penguin.

Garfinkel, H. (2002). *Ethnomethodology's program: Working out Durkheim's aphorism.* Blue Ridge Summit, PA: Rowman and Littlefield.

Garman, M. (1990). *Psycholinguistics.* Cambridge, UK: Cambridge University Press.

Garson, G. D. (2002, June). *Statnotes: An online textbook* [Online]. Retrieved from http://www2.chass.ncsu.edu/garson/pa765/statnote.htm.

Gaventa, J. (1980). *Power and powerlessness: Quiescence and rebellion in an Appalachian valley.* Chicago: University of Chicago Press.

Gee, J. P. (1991). A linguistic approach to narrative. *Journal of Narrative and Life History*, *1*, 15-39.

Geertz, C. (1973). *The interpretation of cultures.* New York: Basic Books.

Geertz, C. (1983). *Local knowledge.* New York: Basic Books.

Geertz, C. (1988). *Works and lives: The anthropologist as author.* Stanford, CA: Stanford University Press.

Geertz, C. (2000). *The interpretation of cultures: Selected essays.* New York: Basic Books.

Geladi, P., & Kowlaski, B. (1986). Partial least square regression: A tutorial. *Analytica Chemica Acta*, *35*, 1-17.

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*, 398-409.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis.* London: Chapman & Hall.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 721-741.

Gensler, H. J. (2002). *Introduction to logic.* London: Routledge Kegan Paul.

Gentle, J. (2003). *Computational statistics.* New York: Springer-Verlag.

George, A. L. (1979). Case studies and theory development: The method of structured, focused comparison. In P. G. Lauren (Ed.), *Diplomacy: New approaches in history, theory, and policy* (pp. 43-68). New York: Free Press.

George, A. L., & McKeown, T. J. (1985). Case studies and theories of organizational decision making. *Advances in Information Processing in Organizations*, *2*, 21-58.

George, A. L., & Smoke, R. (1974). *Deterrence in American foreign policy: Theory and practice.* New York: Columbia University Press.

Georges, R. A., & Jones, M. O. (1995). *Folkloristics: An introduction.* Bloomington: Indiana University Press.

Gephart, R. P. (1986). Deconstructing the defence for quantification in social science: A content analysis of journal articles on the parametric strategy. *Qualitative Sociology*, *9*, 126-144.

Gephart, R. P. (1988). *Ethnostatistics: Qualitative foundations for quantitative research.* Newbury Park, CA: Sage.

Gephart, R. P. (1997). Hazardous measures: An interpretive textual analysis of quantitative sensemaking during crises. *Journal of Organizational Behavior*, *18*, 583-622.

Gerber, A. S., & Green, D. P. (2000). The effects of personal canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *The American Political Science Review*, *94*, 653-664.

Gergen, K. J. (1994) *Realities and relationships.* Cambridge, MA: Harvard University Press.

Gergen, K. J. (1999). *An invitation to social construction.* London: Sage.

Gerhardt, U. (1985). Erzähldaten und Hypothesenkonstruktion. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *37*, 230-256.

Gerhardt, U. (1986). *Patientenkarrieren: Eine idealtypenanalytische Studie.* Frankfurt: Suhrkamp.

Gerhardt, U. (1994). The use of Weberian ideal-type methodology in qualitative data interpretation: An outline for ideal-type analysis. *BMS Bulletin de Methodologie Sociologique* (International Sociological Association, Research Committee 33), *45*, 76-126.

Gerhardt, U. (1999). *Herz und Handlungsrationalität: Eine idealtypen-analytische Studie.* Frankfurt: Suhrkamp.

Gershuny, J. (2000). *Changing times: Work and leisure in postindustrial societies.* Oxford, UK: Oxford University Press.

Ghiselli, E. E., Campbell, J. P., & Zedeck, S. (1981). *Measurement theory for the behavioral sciences.* San Francisco: W. H. Freeman.

Gibbons, J. D. (1988). Truncated data. In S. Kotz, N. L. Johnson, & C. B. Read (Eds.), *Encyclopedia of statistical sciences* (Vol. 9, p. 355). New York: Wiley.

Gibbons, J. D. (1993). *Nonparametric statistics: An introduction.* Newbury Park, CA: Sage.

Gibbons, J. D. (1997). *Nonparametric methods for quantitative analysis* (3rd ed.). Columbus, OH: American Sciences Press.

Gibbons, J. D., & Chakraborti, S. (1992). *Nonparametric statistical inference* (3rd ed.). New York: Marcel Dekker.

Gibson, N., Gibson, G., & Macaulay, A. C. (2001). Community-based research: Negotiating agendas and evaluating outcomes. In J. Morse, J. Swanson, & A. J. Kuzel (Eds.), *The nature of qualitative evidence* (pp. 160-182). Thousand Oaks, CA: Sage.

Giddens, A. (1976). *New rules of sociological method.* London: Hutchinson.

Giddens, A. (1979). *Central problems in social theory.* Berkeley: University of California Press.

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration.* Cambridge, UK: Polity.

Giele, J. Z., & Elder, G. H. (1998). *Methods of life course research: Qualitative and quantitative approaches.* Thousand Oaks, CA: Sage.

Gifi, A. (1990). *Nonlinear multivariate analysis.* New York: Wiley.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650-669.

Gilbert, G. N., & Mulkay, M. (1984). *Opening Pandora's box: A sociological analysis of scientists' discourse.* Cambridge, UK: Cambridge University Press.

Gilbert, N., & Troitzsch, K. G. (1999). *Simulation for the social scientist*. Milton Keynes, UK: Open University Press.

Gill, J. (2000). *Generalized linear models: A unified approach*. Thousand Oaks, CA: Sage.

Gill, J. (2002). *Bayesian methods: A social and behavioral sciences approach*. London: Chapman & Hall.

Gintis, H. (2000). *Game theory evolving*. Princeton, NJ: Princeton University Press.

Girden, E. R. (1992). *ANOVA: Repeated measures* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-084). Newbury Park, CA: Sage.

Glaser, B. (1992). *Emergence vs. forcing: Basics of grounded theory analysis*. Mill Valley, CA: Sociology Press.

Glaser, B. G. (1978). *Theoretical sensitivity*. Mill Valley, CA: Sociology Press.

Glaser, B. G. (2002). Grounded theory and gender relevance. *Health Care for Women International*, *23*, 786-793.

Glaser, B. G., & Strauss, A. L. (1967). *Discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Glass, G. V., & Hopkins, K. D. (1984). *Statistical methods in education and psycho logy* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in education and psycho logy* (3rd ed.). Boston: Allyn & Bacon.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the analysis of variance and convariance. *Review of Educational Research*, *42*, 237-288.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Research*, *5*, 3-8.

Glass, G. V. (1977). Integrating findings. *Review of Research in Education*, *5*, 351-379.

Glass, G. V., & Smith, M. L. (1978). Meta-analysis of research on the relationship of class size and achievement. *Educational Evaluation and Policy Analysis*, *1*, 2-16.

Glenn, N. D. (2003). Distinguishing age, period, and cohort effects. In J. Mortimer & M. Shanahan (Eds.), *Handbook of the life course* (pp. 465-476). New York: Kluwer Academic/Plenum.

Gluck, S. B., & Patai, D. (Eds.). (1991). *Women's words: The feminist practice of oral history*. London: Routledge.

Glymour, C. N. (1980). *Theory and evidence*. Princeton, NJ: Princeton University Press.

Goffman, E. (1959). *The presentation of self in everyday life*. Garden City, NY: Doubleday.

Goffman, E. (1983). The interaction order. *American Sociological Review*, *48*, 1-17.

Gold, R. L. (1958). Roles in sociological field observation. *Social Forces*, *36*, 217-223.

Gold, R. L. (1969). Roles in sociological field observations. In G. J. McCall & J. L. Simmons (Eds.), *Issues in participant observation* (pp. 30-38). Reading, MA: Addison-Wesley.

Goldberger, A. S. (1964). *Econometric theory*. New York: John Wiley & Sons.

Goldmann, L. (1981). *Method in the sociology of literature* (W. Boelhower, Trans. & Ed.). Oxford, UK: Basil Blackwell.

Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Griffin.

Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Hodder Arnold.

Golub, G. H., & Van Loan, G. F. (1996). *Matrix computations* (3rd ed.). Baltimore: Johns Hopkins University Press.

Gomm, R., Hammersley, M., & Foster, P. (2000). Case study and generalisation. In R. Gomm, M. Hammersley, & P. Foster (Eds.), *Case study method: Key issues, key texts*. London: Sage.

Gomm, R., Hammersley, M., & Foster, P. (Eds.). (2000). *Case study method*. London: Sage.

Good, I. J. (1988). The interface between statistics and philosophy of science. *Statistical Science*, *3*, 386-397.

Good, P. (1994). *Permutation tests: A practical guide to resampling methods for testing hypotheses*. New York: Springer.

Goode, W. J. (1978). *The celebration of heroes: Prestige as a control system*. Berkeley: University of California Press.

Goodenough, W. (1957). Cultural anthropology and linguistics. In P. L. Garvin (Ed.), *Report of the 7th Annual Roundtable on Linguistics and Language Study* (pp. 167-173). Washington, DC: Georgetown University Press.

Goodman, A., Johnson, P., & Webb, S. (1997). *Inequality in the UK.* Oxford, UK: Oxford University Press.

Goodman, L. (1953). Ecological regression and the behavior of individuals. *American Sociological Review*, *18*, 663-664.

Goodman, L. A. (1961). Statistical methods for the moverstayer model. *Journal of the American Statistical Association*, *56*, 841-868.

Goodman, L. A. (1973). Causal analysis of data from panel studies and other kinds of surveys. *American Journal of Sociology*, *78*, 1135-1191.

Goodman, L. A. (1973). The analysis of multidimensional contingency tables when some variables are posterior to others: A modified path analysis approach. *Biometrika*, *60*, 179-192.

Goodman, L. A. (1974). The analysis of systems of qualitative variables when some of the variables are unobservable: I. A modified latent structure approach. *American Journal of Sociology*, *79*, 1179-1259.

Goodman, L. A. (1978). *Analyzing qualitative/categorical data.* Cambridge, MA: Abt.

Goodman, L. A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, *74*, 537-552.

Goodman, L. A. (1984). *The analysis of cross-classified data having ordered categories.* Cambridge, MA: Harvard University Press.

Goodman, L. A. (1986). Some useful extensions of the usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables. *International Statistical Review*, *54*, 243-309.

Goodman, L. A., & Hout, M. (1998). Understanding the Goodman-Hout approach to the analysis of differences in association and some related comments. In A. E. Raftery (Ed.), *Sociological methodology* (pp. 249-261). Washington, DC: American Sociological Association.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classification. *Journal of the American Statistical Association*, *49*, 732-764.

Goodman, S. N. (1999). Toward evidence-based medical statistics: 1. The *p* value fallacy. *Annals of Internal Medicine*, *130*, 995-1004.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers.* New York: Academic Press.

Gopaul-McNicol, S.-A., & Armour-Thomas, E. (2002). *Assessment and culture: Psychological tests with minority populations.* San Diego: Academic Press.

Gottman, J. M. (1981). *Time-series analysis: A comprehensive introduction for social scientists.* Cambridge, UK: Cambridge University Press.

Gottschalk, L., Kluckhohn, C., & Angell, R. (1945). *The use of personal documents in history, anthropology, and sociology.* New York: Social Science Research Council.

Gouldner, A. V. (1973). *For sociology.* Harmondsworth, UK: Penguin.

Gouldner, A. W. (1962). Anti-minotaur: The myth of a valuefree sociology. *Social Problems*, *9*, 199-213.

Gower, J. C., & Hand, D. J. (1996). *Biplots.* London: Chapman & Hall.

Gower, J. C., & Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification*, *5*, 5-48.

Graham, J. M., Guthrie, A. C., & Thompson, B. (2003). Consequences of not interpreting structure coefficients in published CFA research: A reminder. *Structural Equation Modeling*, *10*, 142-153.

Grammer, K., Fink, B., & Renninger, L. (2002). Dynamic systems and inferential information processing in human communication. *Neuroendocrinology Letters*, *23*(Suppl. 4), 15-22.

Gramsci, A. (1971). *Selections from the prison notebooks.* London: Lawrence and Wishart.

Granger, C. W. J. (1964). *Spectral analysis of economic time series.* Princeton, NJ: Princeton University Press.

Granger, C. W. J. (1969). Investigating causal

relations by econometric models and cross-spectral methods. *Econometrica*, *37*, 24-36.

Granger, C. W. J. (1981). Some properties of time series data and their use in econometric models specification. *Journal of Econometrics*, *16*, 121-130.

Granger, C. W. J. (1983). *Co-integrated variables and error-correcting models.* Discussion Paper No. 1983-13, University of California, San Diego.

Granger, C. W. J. (1989). *Forecasting in business and economics.* Boston: Academic Press.

Granger, C. W. J., & Newbold, P. (1986). *Forecasting economic time series* (2nd ed.). New York: Academic Press.

Gravetter, F. J., & Wallnau, L. B. (2000). *Statistics for the behavioral sciences* (5th ed.). Belmont, CA: Wadsworth.

Gray, J., & Reuter, A. (1992). *Transaction processing: Concepts and techniques.* San Francisco: Morgan Kaufmann.

Green, D. P., & Gerber, A. S. (2002). Reclaiming the experimental tradition in political science. In I. Katznelson & H. V. Milner (Eds.), *Political science: State of the discipline* (3rd ed., pp. 805-832). New York: W.W. Norton.

Green, D. P., Gerber, A. S., & De Boef, S. L. (1999). Tracking opinion over time: A method for reducing sampling error. *Public Opinion Quarterly*, *63*, 178-192.

Green, P. (1978). *Analyzing multivariate data.* Hinsdale, IL: Dryden.

Green, W. H. (2000). *Econometric analysis* (4th ed.). Upper Saddle River: Prentice Hall.

Greenacre, M. J. (1984). *Theory and method of correspondence analysis.* London: Academic Press.

Greenacre, M. J. (1993). *Correspondence analysis in practice.* London: Academic Press.

Greenacre, M. J., & Blasius, J. (1994). *Correspondence analysis in the social sciences.* London: Academic Press.

Greenbaum, T. L. (1998). *The handbook for focus group research.* London: Sage.

Greene, R. L. (2000). *The MMPI-2: An interpretive manual.* Boston: Allyn & Bacon.

Greene, W. H. (1993). *Econometric analysis* (2nd ed.). New York: Macmillan.

Greene, W. H. (1997). *Econometric analysis* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Greene, W. H. (2000). *Econometric analysis* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

Greene, W. H. (2003). *Econometric analysis* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Greenfield, P. (1997). You can't take it with you: Why ability assessments don't cross cultures. *American Psychologist*, *52*(10), 1115-1124.

Greenhouse, G. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *55*, 431-433.

Greenland, S. (2003). Quantifying biases in causal models. *Epidemiology*, *14*, 300-306.

Greenland, S., & Brumback, B. A. (2002). An overview of relations among causal modelling methods. *International Journal of Epidemiology*, *31*, 1030-1037.

Greenland, S., & Robins, J. M. (1986). Identifiability, exchangeability, and epidemiological confounding. *International Journal of Epidemiology*, *15*, 413-419.

Greenland, S., Robins, J. M., & Pearl, J. (1999). Confounding and collapsibility in causal inference. *Statistical Science*, *14*, 29-46.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, *109*, 3-25.

Greenwood, D., & Levin, M. (1998). *Introduction to action research: Social research for social change.* Thousand Oaks, CA: Sage.

Greimas, A. J., & Courtés, J. (1982). *Semiotics and language: An analytical dictionary.* Bloomington: Indiana University Press.

Griffin, L., Caplinger, C., Lively, K., Malcom, N. L., McDaniel, D., & Nelsen, C. (1997). Comparative-historical analysis and scientific inference: Disfranchisement in the U.S. South as a test case. *Historical Methods*, *30*, 13-27.

Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica*, *25*, 501-522.

Grinyer, A. (2002). The anonymity of research participants: Assumptions, ethics, and practicalities.

Social Research Update, *36*, 1-4.

Grofman, B., & Davidson, C. (1992). *Controversies in minority voting: The voting rights act in perspective.* Washington, DC: Brookings Institution.

Gross, D., & Harris, C. M. (1998). *Fundamentals of queueing theory* (3rd ed.). New York: Wiley.

Groves, E. R., & Ogburn, W. F. (1928). *American marriage and family relationships.* New York: Holt.

Groves, R. M. (1989). *Survey errors and survey costs.* New York: Wiley.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys.* New York: Wiley.

Groves, R. M., Dillman, D. A., Eltinge, J. L., & Little, R. J. A. (Eds.). (2001). *Survey nonresponse.* New York: Wiley.

Guba, E. G. (1978). *Toward a methodology of naturalistic inquiry in educational evaluation.* CSE Monograph Series in Evaluation, No. 8. Los Angeles: Center for the Study of Evaluation, University of California, Los Angeles.

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communications and Technology Journal*, *29*, 75-92.

Guba, E. G., & Lincoln, Y. S. (1985). *Naturalistic inquiry.* Beverly Hills, CA: Sage.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation.* Newbury Park, CA: Sage.

Guba, E., & Lincoln, Y. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105-117). Thousand Oaks, CA: Sage.

Gubrium, J. F., & Buckholdt, D. R. (1979). Production of hard data in human service organizations. *Pacific Sociological Review*, *22*, 115-136.

Gubrium, J. F., & Holstein, J. A. (1997). *The new language of qualitative method.* New York: Oxford University Press.

Gubrium, J. F., & Holstein, J. A. (2002). From the individual interview to the interview society. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research: Context and method* (pp. 3-32). Thousand Oaks, CA: Sage.

Gujarati, D. (1995). *Basic econometrics* (3rd ed.). New York: McGraw-Hill.

Gujarati, D. (2002). *Basic econometrics* (4th ed.). New York: McGraw-Hill.

Gulliksen, H. (1950). *The theory of mental tests.* New York: Wiley.

Guttman, L. (1941). An outline for the statistical theory of prediction. In P. Horst (with collaboration of P. Wallin & L. Guttman) (Ed.), *The prediction of personal adjustment* (Bulletin 48, Supplementary Study B-1, pp. 253-318). New York: Social Science Research Council.

Guttman, L. (1944). A basis for scaling quantitative data. *American Sociological Review*, *9*, 139-150.

Guttman, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika*, *18*, 277-296.

Guttman, L. (1954). Some necessary conditions for commonfactor analysis. *Psychometrika*, *19*, 149-161.

Guttman, L. (1956). "Best possible" systematic estimates of communalities. *Psychometrika*, *21*, 273-285.

Haack, S. (1993). *Evidence and inquiry: Toward a reconstruction of epistemology.* Oxford, UK: Blackwell.

Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica*, *12*(Suppl.), preface, p. iii.

Haberman, S. J. (1979). *Analysis of qualitative data: Vol. 2. New developments.* New York: Academic Press.

Hacking, I. (1965). *The logic of statistical inference.* Cambridge, UK: Cambridge University Press.

Hacking, I. (2001). *An introduction to probability and inductive logic.* Cambridge, UK: Cambridge University Press.

Hagenaars, J. A. (1990). *Categorical longitudinal data: Loglinear analysis of panel, trend and cohort data.* Newbury Park, CA: Sage.

Hagenaars, J. A. (1993). *Loglinear models with latent variables.* Newbury Park, CA: Sage.

Hagenaars, J. A. (1998). Categorical causal modeling: Latent class analysis and directed log-linear models with latent variables. *Sociological Methods and Research*, *26*, 436-486.

Hagenaars, J. A., & McCutcheon, A. L. (2002).

Applied latent class analysis. Cambridge, UK: Cambridge University Press.

Hagle, T. M., & Mitchell, G. E., Ⅱ. (1992). Goodness-of-fit measures for probit and logit. *American Journal of Political Science*, *36*, 762-784.

Haight, B. K., Michel, Y., & Hendrix, S. (1998). Life review: Preventing despair in newly relocated nursing home residents: Short- and long-term effects. *International Journal of Aging and Human Development*, *47*(2), 119-142.

Hajek, J., & Sidák, Z. (1967). *Theory of rank tests.* New York: Academic Press.

Hakim, C. (1982). *Secondary analysis of social research.* London: George Allen and Unwin.

Haladyna, T. M. (1994). *Developing and validating multiplechoice test items.* Hillsdale, NJ: Lawrence Erlbaum.

Hald, A. (1952). *Statistical theory with engineering applications.* New York: John Wiley.

Hald, A. (1990). *A history of probability and statistics and their applications before 1750.* New York: John Wiley and Sons.

Halfpenny, P. (1982). *Positivism and sociology: Explaining social life.* London: Allen & Unwin.

Hall, E. T. (1974). *Handbook for proxemic research.* Washington, DC: Society for the Anthropology of Visual Communication.

Hall, S. (1974). The television discourse—Encoding and decoding. In *Education and culture* (Vol. 35, pp. 8-14). Paris: United Nations Educational, Social, and Cultural Organization.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Hamilton, D. (1980). Some contrasting assumptions about case study research and survey analysis. In H. Simons (Ed.), *Towards a science of the singular: Essays about case study in educational research and evaluation* (pp. 78-92). Norwich, UK: Centre for Applied Research in Education University of East Anglia.

Hamilton, J. D. (1994). *Time series analysis.* Princeton, NJ: Princeton University Press.

Hamilton, L. C. (1990). *Modern data analysis: A first course in applied statistics.* Pacific Grove, CA: Brooks/Cole.

Hamilton, L. C. (1992). *Regression with graphics: A second course in applied statistics.* Pacific Grove, CA: Brooks/Cole.

Hamlyn, D. W. (1967). Empiricism. In P. Edwards (Ed.), *The encyclopedia of philosophy* (pp. 499-505). New York: Macmillan.

Hammersley, M. (1989). *The dilemma of qualitative method: Herbert Blumer and the Chicago school of sociology.* London: Routledge and Kegan Paul.

Hammersley, M. (1989). The problem of the concept: Herbert Blumer on the relationship between concepts and data. *Journal of Contemporary Ethnography*, *18*, 133-160.

Hammersley, M. (1990). *Reading ethnographic research.* London: Longman.

Hammersley, M. (1992). *What's wrong with ethnography.* London: Routledge Kegan Paul.

Hammersley, M. (1995). *The politics of social research.* London: Sage.

Hammersley, M. (1995). Theory and evidence in qualitative research. *Quality and Quantity*, *29*, 55-66.

Hammersley, M. (1997). Qualitative data archiving: Some reflections on its prospects and problems. *Sociology*, *31*(1), 131-142.

Hammersley, M. (1999). Sociology, what's it for? A critique of the grand conception. *Sociological Research Online*, *4*(3). Retrieved from http://www.socresonline.org.uk/ socresonline/4/3/hammersley.html.

Hammersley, M. (2000). *Taking sides in social research.* London: Routledge.

Hammersley, M. (2001). On "systematic" reviews of research literatures: A "narrative" reply to Evans and Benefield. *British Educational Research Journal*, *27*(5), 543-554.

Hammersley, M., & Atkinson, P. (1995). *Ethnography: Principles in practice.* London: Routledge.

Hammond, M., Howarth, J., & Keat, R. (1991). *Understanding phenomenology.* Oxford, UK: Blackwell.

Handcock, M. S., & Morris, M. (1999). *Relative distribution methods in the social sciences.* New York: Springer-Verlag.

Haney, C., Banks, C., & Zimbardo, P. (1973).

Interpersonal dynamics in a simulated prison. *International Journal of Criminology and Penology*, *1*, 69-97.

Hansen, M. H., Hurwitz, W. N., & Madow, W. G. (1953). *Sample survey methods and theory.* New York: John Wiley & Sons.

Hansen, P. R., & Johansen, S. (1998). *Workbook on cointegration.* Oxford, UK: Oxford University Press.

Hantrais, L., & Mangen, S. (Eds.). (1996). *Cross-national research methods in the social sciences.* London: Pinter.

Harary, F., Norman, D., & Cartwright, D. (1965). *Structural models for directed graphs.* New York: Free Press.

Hardin, J. W., & Hilbe, J. M. (2003). *Generalized estimating equations.* London: Chapman & Hall.

Harding, S. (1991). *Whose science? Whose knowledge?* Milton Keynes, UK: Open University Press.

Harding, S. (1992). After the neutrality ideal: Science, politics, and "strong objectivity." *Social Research*, *59*, 568-587.

Harding, S. (Ed.). (1987). *Feminism and methodology.* Milton Keynes, UK: Open University Press.

Hardy, M. A. (1993). *Regression with dummy variables* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-093). Newbury Park, CA: Sage.

Hardy, M. A., & Reynolds, J. (2003). Incorporating categorical information into regression models: the utility of dummy variables. In M. A. Hardy & A. Bryman (Eds.), *Handbook of data analysis.* London: Sage.

Hare, R. D. (1966). Temporal gradient of fear arousal in psychopaths. *Journal of Abnormal and Social Psychology*, *70*, 442-445.

Harper, D. (2001). *Changing works: Visions of a lost agriculture.* Chicago: University of Chicago Press.

Harper, D. (2002). Talking about pictures: A case for photo elicitation. *Visual Studies*, *17*(1), 13-26.

Harré, R. (1961). *Theories and things.* London: Sheed & Ward.

Harré, R. (1986). *Varieties of realism.* Oxford, UK: Blackwell.

Harré, R., & Secord, P. F. (1972). *The explanation of social behaviour.* Oxford, UK: Basil Blackwell.

Harris, C. W. (1962). Some Rao-Guttman relationships. *Psychometrika*, *27*, 247-263.

Harris, J. A. (1913). On the calculation of intraclass and interclass coefficients of correlation from class moments when the number of possible combinations is large. *Biometrika*, *9*, 446-472.

Harris, M. (1976). History and significance of the emic/etic distinction. *Annual Review of Anthropology*, *5*, 329-350.

Hart, C. (1998). *Doing a literature review: Releasing the social science research imagination.* London: Sage.

Hartigan, J. A., & Kleiner, B. (1981). Mosaics for contingency tables. In W. F. Eddy (Ed.), *Computer science and statistics: Proceedings of the 13th Symposium on the Interface* (pp. 268-273). New York: Springer-Verlag.

Hartsock, N. C. M. (1998). *The feminist standpoint revisited and other essays.* Boulder, CO: Westview.

Hartwig, F., & Dearing, B. E. (1979). *Exploratory data analysis* (Sage University Papers on Quantitative Applications in the Social Sciences). Beverly Hills, CA: Sage.

Harvey, A. (1990). *The econometric analysis of time series* (2nd ed.). Cambridge: MIT Press.

Harvey, A. C. (1990). *Forecasting, structural time series models, and the Kalman filter.* New York: Cambridge University Press.

Harvey, A. S. (1999). Guidelines. In W. E. Pentland, A. S. Harvey, M. P. Lawton, & M. A. McColl (Eds.), *Time use research in the social sciences* (pp. 19-45). New York: Kluwer.

Harvey, D. (1973). *Social justice and the city.* Baltimore: Johns Hopkins University Press.

Harvey, D. (1989). *The condition of postmodernity.* Oxford, UK: Basil Blackwell.

Harvey, L. (1987). *Myths of the Chicago school of sociology.* Aldershot, England: Avebury.

Hastie, T. J. (1992). Generalized additive models. In J. M. Chambers & T. J. Hastie (Eds.), *Statistical models in S* (pp. 249-307). Pacific Grove, CA: Wadsworth and Brooks/Cole.

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models.* New York: Chapman &

Hall.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning*. New York: Springer-Verlag.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains. *Biometrika*, *57*, 97-109.

Hauser, R. M. (1978). A structural model of the mobility table. *Social Forces*, *56*, 919-953.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, *46*, 1251-1271.

Hausman, J., & McFadden, D. (1984). Specification tests for the multinomial logit model. *Econometrica*, *52*, 1219-1240.

Hayano, D. (1979). Auto-ethnography: Paradigms, problems and prospects. *Human Organization*, *38* (1), 99-104.

Hayes, S. C., White, D., & Bissett, R. T. (1998). Protocol analysis and the "silent dog" method of analyzing the impact of self-generated rules. *Analysis of Verbal Behavior*, *15*, 57-63.

Hays, W. L. (1972). *Statistics for the social sciences*. New York: Holt.

Hays, W. L. (1988). *Statistics*. New York: Holt, Rinehart and Winston.

Hays, W. L. (1994). *Statistics* (5th ed.). Fort Worth, TX: Harcourt Brace.

Headland, T. N., Pike, K. L., & Harris, M. (1990). *Emics and etics: The insider/outsider debate*. Newbury Park, CA: Sage.

Healey, J. F. (1995). *Statistics: A tool for social research* (3rd ed.). Belmont, CA.: Wadsworth.

Heaton, J. (1998). Secondary analysis of qualitative data [Online]. *Social Research Update*, *22*. Retrieved April 28, 2002, from http://www.soc.surrey.ac.uk/sru/SRU22.html.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*, 153-161.

Heckman, J. J., & Singer, B. (1982). Population heterogeneity in demographic models. In K. Land & A. Rogers (Eds.), *Multidimensional mathematical demography* (pp. 567-599). New York: Academic Press.

Heckman, J. J., & Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, *52*(2), 271-320.

Heckman, J. J., & Smith, J. A. (1995). Assessing the case for social experiments. *Journal of Economic Perspectives*, *9*, 85-110.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hedstrom, P., & Swedberg, R. (1998). *Social mechanisms: An analytical approach to social theory*. Cambridge, UK: Cambridge University Press.

Heider, K. G. (1976). *Ethnographic film*. Austin: University of Texas Press.

Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences*. Thousand Oaks, CA: Sage.

Heise, D. (1971). The semantic differential and attitude research. In G. F. Summers (Ed.), *Attitude measurement* (pp. 235-253). Chicago: Rand McNally.

Heise, D. R. (1975). *Causal analysis*. New York: Wiley.

Heise, D. R. (1987). Affect Control Theory: Concepts and model. *Journal of Mathematical Sociology*, *13*, 1-33.

Hektner, J. M., & Csikszentmihalyi, M. (2002). The experience sampling method: Measuring the context and the content of lives. In R. B. Bechtel & A. Churchman (Eds.), *Handbook of environmental psychology* (pp. 233-243). New York: John Wiley & Sons.

Helland, I. S. (1990). PLS regression and statistical models. *Scandinavian Journal of Statistics*, *17*, 97-114.

Hempel, C. (1965). *Aspects of scientific explanation, and other essays in the philosophy of science*. New York: Free Press.

Hempel, C. (1965). Confirmation, induction, and rational belief. In C. Hempel (Ed.), *Aspects of scientific explanation*. New York: Free Press.

Hempel, C. G. (1966). *Philosophy of natural science*. Englewood Cliffs, NJ: Prentice Hall.

Hendrick, T., Bickman, L., & Rog, D. J. (1993). *Applied research design: A practical guide*. Newbury Park, CA: Sage.

Hendry, D. F. (1995). *Dynamic econometrics*. New

York: Oxford University Press.

Henkel, R. E. (1976). *Tests of significance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-004). Beverly Hills, CA: Sage.

Henle, M., & Hubble, M. B. (1938). Egocentricity in adult conversation. *Journal of Social Psychology*, *9*, 227-234.

Henry, G. T. (1990). *Practical sampling.* Newbury Park, CA: Sage.

Henry, G. T., & Gordon, C. S. (2001). Tracking issue attention: Specifying the dynamics of the public agenda. *Public Opinion Quarterly*, *65*(2), 157-177.

Hepburn, A. (2000). On the alleged incompatibility between feminism and relativism. *Feminism and Psychology*, *10*, 91-106.

Heritage, J. (1984). *Garfinkel and ethnomethodology.* Cambridge, UK: Polity.

Herrnstein, R., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: Free Press.

Hess, I. (1985). *Sampling for social research surveys 1947-1980.* Ann Arbor, MI: Institute for Social Research.

Hewitt, J. (1989). *Dilemmas of the American self.* Philadelphia: Temple University Press.

Heyde, C. C., & Seneta, E. (Eds.). (2001). *Statisticians of the centuries.* New York: Springer.

Hibbs, D. A. (1982). The dynamics of political support for American presidents among occupational and partisan groups. *American Journal of Political Science*, *26*, 312-332.

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, *13*, 141-154.

Hill, A. B. (1953). Observation and experiment. *New England Journal of Medicine*, *248*, 995-1001.

Hill, M. M. (1993). *Archival strategies and techniques* (Qualitative Research Methods Series No. 31). Newbury Park, CA: Sage.

Hill, M. S. (1992). *The panel study of income dynamics: A user's guide.* Newbury Park, CA: Sage.

Hilsum, S., & Cane, B. (1971). *The teacher's day.* Windsor, UK: National Foundation for Educational Research.

Himmelfarb, G. (1987). *The new history and the old.* Cambridge, MA: Harvard University Press.

Hinchman, L. P., & Hinchman, S. K. (Eds.). (1997). *Memory, identity, community: The idea of narrative in the human sciences.* Albany: State University of New York Press.

Hinds, P. S., Vogel, R. J. . & Clarke-Steffen, L. (1997). The possibilities and pitfalls of doing a secondary analysis of a qualitative data set. *Qualitative Health Research*, *7*, 408-424.

Hine, C. (2000). *Virtual ethnography.* London: Sage.

Hinton, P. R. (1996). *Statistics explained: A guide for social science students.* London: Routledge.

Hirsch, M. W., & Smale, S. (1974). *Differential equations, dynamical systems, and linear algebra.* New York: Academic Press.

Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (Eds.). (1983). *Understanding robust and exploratory data analysis.* New York: Wiley.

Hobcraft, J., Menken, J., & Preston, S. (1982). Age, period, and cohort effects in demography: A review. *Population Index*, *42*, 4-43.

Hobert, J. P., & Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, *91*, 1461-1473.

Hobson-West, P., & Sapsford, R. (2001). *The Middlesbrough Town Centre Study: Final report.* Middlesbrough, UK: University of Teesside, School of Social Sciences.

Hochberg, Y., & Tamhane, A. C. (1987). *Multiple comparison procedures.* New York: John Wiley.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression. *Technometrics*, *12*(1), 55-67, 69-82.

Hoey, M. (2001). *Textual interaction.* London: Routledge.

Hofmann, H. (2000). Exploring categorical data: Interactive mosaic plots. *Metrika*, *51*(1), 11-26.

Hogg, R. V., & Craig, A. T. (1978). *Introduction to mathematical statistics* (4th ed.). New York: Macmillan.

Hogg, R. V., & Tanis, E. A. (1988). *Probability and statistical inference.* New York: Macmillan.

Hollander, M., & Wolfe, D. A. (1973).

Nonparametric statistical inference. New York: Wiley.

Hollander, M., & Wolfe, D. A. (1999). *Nonparametric statistical methods* (2nd ed.). New York: Wiley.

Hollis, M. (1977). *Models of man.* Cambridge, UK: Cambridge University Press.

Hollis, M., & Smith, S. (1990). *Explaining and understanding international relations.* New York: Oxford University Press.

Hollway, W., & Jefferson, T. (2000). *Doing qualitative research differently: Free association, narrative and the interview method.* London: Sage.

Holmes, D. (1976). Debriefing after psychological experiments: Effectiveness of post-experimental desensitizing. *American Psychologist*, *32*, 868-875.

Holmwood, J. (2001). Gender and critical realism: A critique of Sayer. *Sociology*, *35*, 947-965.

Holstein, J. A., & Gubrium, J. F. (1995). *The active interview.* Thousand Oaks, CA: Sage.

Holstein, J. A., & Gubrium, J. F. (2000). *The self we live by: Narrative identity in a postmodern world.* NewYork: Oxford University Press.

Holsti, O. R. (1969). *Content analysis for the social sciences and humanities.* Reading, MA: Addison-Wesley.

Homan, R. (1992). *The ethics of social research.* London: Longman.

Homans, G. C. (1964). Contemporary theory in sociology. In R. E. L. Faris (Ed.), *Handbook of modern sociology* (pp. 951-977). Chicago: Rand McNally.

Homans, G. C. (1967). *The nature of social science.* New York: Harcourt, Brace, & World.

Honneth, A. (1996). *The struggle for recognition: The moral grammar of social conflicts* (J. Anderson, Trans.). Cambridge: MIT Press.

Hopkins, K. D. (1982). The unit of analysis: Group means versus individual observations. *American Educational Research Journal*, *19*(1), 5-18.

Horn, R. (1996). Negotiating research access to organizations. *Psychologist*, *9*(12), 551-554.

Horst, P. (1941). The role of the predictor variables which are independent of the criterion. *Social Science Research Council*, *48*, 431-436.

Horton, M., & Freire, P. (2000). *We make the road by walking* (Bell, B., Gaventa, J., & Peters, J., Eds.). Philadelphia: Temple University Press.

Horton, N. J., & Lipsitz, S. R. (2001). Multiple imputation in practice: Comparison of software packages for regression models with missing variables. *American Statistician*, *55*, 244-254.

Horvitz, D. G., Shah, B. U., & Simmons, W. R. (1967). The unrelated question randomized response model. In E. D. Goldfield (Ed.), *Proceedings of the Social Statistics Section* (pp. 65-72). Washington, DC: American Statistical Association.

Hoskins, J. (1998). *Biographical objects: How things tell the stories of people's lives.* London: Routledge.

Höskuldsson, A. (1988). PLS regression methods. *Journal of Chemometrics*, *2*, 211-228.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, *24*, 417-441, 498-520.

Hougaard, P. (2000). *Analysis of multivariate survival data.* New York: Springer-Verlag.

House, E. R., & Howe, K. R. (1999). *Values in evaluation and social research.* Thousand Oaks, CA: Sage.

Hout, M. (1983). *Mobility tables.* Beverly Hills, CA: Sage.

Howell, D. C. (1999). *Fundamental statistics for the behavioral sciences* (4th ed.). Belmont, CA: Duxbury.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Duxbury, UK: Thomson Learning.

Hox, J. J. (2002). *Multilevel analysis, techniques and applications.* Mahwah, NJ: Lawrence Erlbaum.

Hoyle, R. H. (2000). Confirmatory factor analysis. In H. E. A. Tinsely & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 465-497). New York: Academic Press.

Hsiao, C. (1986). *Analysis of panel data.* Cambridge, UK: Cambridge University Press.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary*

Journal, *6*, 1-55.

Huber, P. (1981). *Robust statistics.* New York: Wiley.

Huberman, A. M., & Miles, M. B. (1994). Data management and analysis methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 428-444). Thousand Oaks, CA: Sage.

Huck, S. (2000). *Reading statistics and research* (3rd ed.). New York: Addison Wesley Longman.

Huff, A. S. (2000). Changes in organizational knowledge production. *Academy of Management Review*, *25*(2), 288-293.

Hughes, J. A. (1990). *The philosophy of social research* (2nd ed.). Harlow, UK: Longman.

Hughes, K., MacKintosh, A. M., Hastings, G., Wheeler, C., Watson, J., & Inglis, J. (1997). Young people, alcohol, and designer drinks: A quantitative and qualitative study. *British Medical Journal*, *314*, 414-418.

Hughes, R. (1998). Considering the vignette technique and its application to a study of drug injecting and HIV risk and safer behaviour. *Sociology of Health and Illness*, *20*, 381-400.

Hujala, E. (1998). Problems and challenges in cross-cultural research. *Acta Universitatis Ouluensis*, *35*, 19-31.

Hume, D. (1786). *Treatise on human nature.* Oxford, UK: Claridon.

Humphreys, L. (1975). *Tearoom trade: Impersonal sex in public places.* Chicago: Aldine.

Hunt, J. C. (1989). *Psychoanalytic aspects of fieldwork* (University Paper Series on Qualitative Research Methods 18). Newbury Park, CA: Sage.

Hunter, J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin*, *86*, 721-735.

Husserl, E. (1970). *Logical investigations* (Vols. 1-2). London: Routledge Kegan Paul.

Hutcheson, G., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models.* London: Sage.

Huynh, H., & Feldt, L. S. (1970). Conditions under which mean square ratios in repeated measurements designs have exact *F*-distributions. *Journal of the American Statistical Association*, *65*, 1582-1589.

Hyde, R. (2003). *The art of assembly language.* San Francisco: No Starch Press.

Hyman, H. (1955). *Survey design and analysis.* New York: Free Press.

Hyman, H. H. (1972). *Secondary analysis of sample surveys.* Glencoe, IL: Free Press.

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis.* New York: Wiley.

ICPSR. (2002). *Guide to social science data preparation and archiving* [Online]. Available: http://www. ifdo. org/archiving-distribution/datprep-archiving-bfr.htm.

IFDO. (n. d.). [Online]. Available: http://www.ifdo.org/.

Imber, J. B. (Ed.). (2001). Symposium: Population politics. *Society*, *39*, 3-53.

Inkeles, A., & Sasaki, M. (Eds.). (1996). *Comparing nations and cultures.* Englewood Cliffs, NJ: Prentice Hall.

International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDRB). (1992). *Demographic surveillance system—Matlab: Registration of demographic events—1985* (Scientific Report 68). Dhaka, Bangladesh: Author.

Inter-University Consortium for Political and Social Research. (2002). *Guide to social science data preparation and archiving* [Online]. Supported by the Robert Wood Johnson Foundation. Available: www.ICPSR.umich.edu.

Iversen, G. R. (1973). Recovering individual data in the presence of group and individual effects. *American Journal of Sociology*, *79*, 420-434.

Iversen, G. R. (1984). *Bayesian statistical inference* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-043). Beverly Hills, CA: Sage.

Iversen, G. R. (1991). *Contextual analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-081). Newbury Park, CA: Sage.

Iversen, G. R., & Norpoth, H. (1987). *Analysis of variance* (2nd ed., Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-001). Newbury Park, CA: Sage.

Iversen, G., & Gergen, M. (1997). *Statistics: The*

conceptual approach. New York: Springer-Verlag.

Iyengar, S., Peters, M. E., & Kinder, D. R. (1982). Experimental demonstrations of the "not-so-minimal" consequences of television news programs. *The American Journal of Political Science*, *4*, 848-858.

Jaccard, J. (1998). *Interaction effects in factorial analysis of variance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-118). Thousand Oaks, CA: Sage.

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression* (2nd ed., Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-72). Thousand Oaks, CA: Sage.

Jaccard, J., Turrisi, R, & Wan, C. K.. (1990). *Interaction effects in multiple regression.* Newbury Park, CA: Sage.

Jackman, R. W. (1973). On the relationship of economic development to political performance. *American Journal of Political Science*, *17*, 611-621.

Jackson, D. N. (2002). The constructs in people's heads. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 3-18). Mahwah, NJ: Lawrence Erlbaum.

Jackson, J. E. (1991). *A user's guide to principal component analysis.* New York: John Wiley.

Jackson, S., & Brashers, D. E. (1994). *Random Factors in ANOVA* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-098). Thousand Oaks, CA: Sage.

Jacobs, H. A. (1861). *Incidents in the life of a slave girl: Written by herself.* Boston.

Jacoby, L. L., Debner, J. A., & Hay, J. F. (2001). Proactive interference, accessibility bias, and process dissociations: Valid subject reports of memory. *Journal of Experimental Psychology: Memory, Learning, and Cognition*, *27*, 686-700.

Jacoby, W. G. (1991). *Data theory and dimensional analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-078). Newbury Park, CA: Sage.

Jacoby, W. G. (1997). *Statistical graphics for univariate and bivariate data.* Thousand Oaks, CA: Sage.

Jaeger, R. M. (1993). *Statistics: A spectator sport* (2nd ed.). Newbury Park, CA: Sage.

Jahoda, G. (1977). In pursuit of the emic-etic distinction: Can we ever capture it? In Y. H. Poortinga (Ed.), *Basic problems in cross-cultural research* (pp. 55-63). Amsterdam: Swets & Zeitlinger.

James, J. B., & Sørensen, A. (2000, December). Archiving longitudinal data for future research: Why qualitative data add to a study's usefulness. *Forum Qualitative Sozialforschung* [Forum: Qualitative Social Research] [Online journal], *1* (3). Available: http://qualitativeresearch.net/fqs/fqs-eng.htm.

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, *69*, 85-98.

James, W. (1952). *The varieties of religious experience: A study in human nature.* London: Longmans, Green & Co. (Original work published 1902)

Jameson, F. (1983). Postmodernism and consumer society. In H. Foster (Ed.), *Postmodern culture.* London: Pluto.

Jasso, G. (1988). Principles of theoretical analysis. *Sociological Theory*, *6*, 1-20.

Jasso, G. (1990). Methods for the theoretical and empirical analysis of comparison processes. *Sociological Methodology*, *20*, 369-419.

Jasso, G. (2001). Formal theory. In J. H. Turner (Ed.), *Handbook of sociological theory* (pp. 37-68). New York: Kluwer Academic/Plenum.

Jasso, G. (2001). Rule-finding about rule-making: Comparison processes and the making of norms. In M. Hechter & K.-D. Opp (Eds.), *Social norms* (pp. 348-393). New York: Russell Sage.

Jasso, G. (2001). Comparison theory. In J. H. Turner (Ed.), *Handbook of sociological theory* (pp. 669-698). New York: Kluwer Academic/Plenum.

Jasso, G. (2001). Studying status: An integrated framework. *American Sociological Review*, *66*, 96-124.

Jasso, G. (2002). Seven secrets for doing theory. In J. Berger & M. Zelditch (Eds.), *New directions in contemporary sociological theory* (pp. 317-342).

Boulder, CO: Rowan & Littlefield.

Jasso, G. (in press). The tripartite structure of social science analysis. *Sociological Theory.*

Jasso, G., & Rossi, P. H. (1977). Distributive justice and earned income. *American Sociological Review*, *42*, 639-651.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Clarendon.

Jenkins, G. M., & Watts, D. G. (1968). *Spectral analysis and its applications.* San Francisco: Holden-Day.

Jenkins, R. (1984). Bringing it all back home: An anthropologist in Belfast. In C. Bell & H. Roberts (Eds.), *Social researching: Politics, problems, practice* (pp. 147-164). London: Routledge and Kegan Paul.

JMP Introductory Guide. (2000). Cary, NC: SAS Institute.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, *12*, 231-254.

Johansen, S. (1991). Estimation and hypotheses testing of cointegrating vectors in Gaussian vector autoregressive models. *Econometrica*, *59*, 1551-1580.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models.* Oxford, UK: Oxford University Press.

John, K. E. (1989). The polls—A report. *Public Opinion Quarterly*, *53*, 590-605.

John, N. R., & Draper, J.A. (1980). An alternative family of transformations. *Applied Statistics*, *29* (2), 190-197.

Johnson, C. (1982). Risks in the publication of fieldwork. In J. E. Sieber (Ed.), *The ethics of social research: Fieldwork, regulation and publication* (pp. 71-92). New York: Springer-Verlag.

Johnson, C. (1994). Gender, legitimate authority, and leadersubordinate conversations. *American Sociological Review*, *59*, 122-135.

Johnson, J. C. (1990). *Selecting ethnographic informants.* Newbury Park, CA: Sage.

Johnson, J. C., & Weller, S. C. (2002). Elicitation techniques for interviewing. In J. F. Gubrium & J. A. Holstein (Eds.), *The handbook of interview research.* Thousand Oaks, CA: Sage.

Johnson, N. L., & Kotz, S. (1969). *Discrete distributions.* Boston: Houghton-Mifflin.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate distributions, Vol. 1.* New York: Wiley.

Johnson, R. A., & Bhattacharyya, G. K. (2001). *Statistics: Principles and methods* (4th ed.). New York: Wiley.

Johnson, R. A., & Wichern, D.W. (2002). *Applied multivariate statistical analysis.* Upper Saddle River, NJ: Prentice Hall.

Johnson, T., Dandeker, C., & Ashworth, C. (1984). *The structure of social theory.* London: Macmillan.

Johnston, J., & Dinardo, J. (1997). *Econometric methods* (4th ed.). New York: McGraw-Hill.

Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards.* Thousand Oaks, CA: Sage.

Joliffe, I. T. (1986). *Principal component analysis.* New York: Springer-Verlag.

Jones, J. M. G., & Hunter, D. (1995). Consensus methods for medical and health services research. *British Medical Journal*, *311*, 376-380.

Jones, K. (1997). Multilevel approaches to modelling contextuality: From nuisance to substance in the analysis of voting beha-viour. In G. P. Westert & R. N. Verhoeff (Eds.), *Places and people: Multilevel modelling in geographical research* (Nederlandse Geografische Studies 227). Utrecht, The Netherlands: The Royal Dutch Geographical Society and Faculty of Geographical Sciences, Utrecht University.

Jones, M. B. (1959). *Simplex theory* (U.S. Naval School of Aviation Medicine Monograph Series No. 3). Pensacola, FL: U.S. Naval School of Aviation Medicine.

Jones, M. O. (1996). *Studying organizational symbolism: What, how, why?* Thousand Oaks, CA: Sage.

Jones, S. R. (1992). Was there a Hawthorne effect? *American Journal of Sociology*, *98*, 451-468.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, *32*, 443-482.

Jöreskog, K. G. (1969). A general approach to

confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*, 183-202.

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85-112). New York: Academic Press.

Jöreskog, K. G. (1979). Statistical models and methods for analysis of longitudinal data. In K. G. Jöreskog & D. Sörbom (Eds.), *Advances in factor analysis and structural equation models*. Cambridge, MA: Abt.

Josselson, R., & Lieblich, A. (Eds.). (1993-1999). *The narrative study of lives* (Vols. 1-6). Newbury Park, CA: Sage.

Josselsyn, R. (Ed.). (1996). *Ethics and process in the narrative study of lives*. Thousand Oaks, CA: Sage.

Joynson, R. B. (1989). *The Burt affair*. London: Routledge.

Judd, C. M., & McClelland, G. H. (1998). Measurement. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1). Boston: McGraw-Hill.

Judd, C. M., Smith, E. R., & Kidder, L. H. (1991). *Research methods in social relations* (6th ed.). Fort Worth, TX: Harcourt Brace Jovanovich.

Judge, G. G., Griffiths, W. E., Hill, R. C., & Lee, T.-C. (1980). *The theory and practice of econometrics* (Wiley Series in Probability and Mathematical Statistics). New York: John Wiley.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H., & Lee, T.-C. (1985). *The theory and practice of econometrics* (2nd ed.). New York: John Wiley.

Judge, G. G., Hill, R. C., Griffiths, W. E., Lütkepohl, H., & Lee, T. (1988). *Introduction to the theory and practice of econometrics* (2nd ed.). New York: John Wiley.

Julian, D. A. (1997). The utilization of the logic model as a system level planning and evaluation device. *Evaluation and Program Planning*, *20*(3), 251-257.

Juster, F. T., & Stafford, F. P. (1985). *Time, goods and wellbeing*. Ann Arbor, MI: Institute for Social Research.

Kalichman, S. C. (1999). *Mandated reporting of suspected child abuse: Ethics, law and policy*. Washington, DC: American Psychological Association.

Kalton, G. (1983). *Introduction to survey sampling* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-035). Beverly Hills, CA: Sage.

Kan, M. (2002). Reinterpreting the Multifactor Leadership Questionnaire. In K. W. Parry & J. R. Meindl (Eds.), *Grounding leadership theory and research: Issues, perspectives and methods* (pp. 159-173). Greenwich, CT: Information Age Publishing.

Kane, M. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice*, *21*(1), 31-41.

Kant, I. (1781). *Kritik der reinen Vernunft* [Critique of pure reason]. Hamburg: Felix Meiner.

Kanuha, V. K. (2000). "Being" native versus "going native": Conducting social work research as an insider. *Social Work*, *45*(5), 439-447.

Kaplan, A. (1943). Content analysis and the theory of signs. *Philosophy of Science*, *10*, 230-247.

Kaplan, A. (1964). *The conduct of inquiry*. New York: Chandler.

Kaplan, D. (1990). Evaluation and modification of covariance structure models: A review and recommendation. *Multivariate Behavioral Research*, *25*, 137-155.

Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Thousand Oaks, CA: Sage.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*, 457-481.

Kashy, D. A., & Kenny, D. A. (2000). The analysis of data from dyads and groups. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 451-477). New York: Cambridge University Press.

Kasprzyk, D., Duncan, G. J., Kalton, G., & Singh, M. P. (1989). *Panel surveys*. New York: John Wiley.

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, *29*(2), 119-127.

Kass, R. E., & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, *91*, 1343-1370.

Kazdin, A. E. (2003). *Research design in clinical psychology, 4th Edition.* Boston: Allyn & Bacon.

Keat, R., & Urry, J. (1975). *Social theory as science.* London: Routledge Kegan Paul.

Keeter, S., Miller, C., Kohut, A., Groves, R. M., & Presser, S. (2000). Consequences of reducing nonresponse in a national telephone survey. *Public Opinion Quarterly*, *64*, 125-148.

Keleman, M., & Bansal, P. (2002). The conventions of management research and their relevance to management practice. *British Journal of Management*, *13*, 97-108.

Kelle, U. (Ed.). (1995). *Computer-aided qualitative data analysis: Theory, methods and practice.* London: Sage.

Kelly, G. (1955). *The psychology of personal constructs.* New York: Norton.

Kendall, G., & Wickham, G. (1999). *Using Foucault's methods.* London: Sage.

Kendall, M. G. (1962). *Rank correlation methods* (3rd ed.). London: Griffin.

Kendall, M. G., & Buckland, W. R. (1971). *A dictionary of statistical terms* (3rd ed.). Edinburgh: Oliver & Boyd.

Kendall, M., & Gibbons, J. D. (1990). *Rank correlation methods* (5th ed.). New York: Oxford University Press.

Kendall, M., & Stuart, A. (1969). *The advanced theory of statistics: Vol. 2. Inference and relationship.* London: Griffin.

Kendall, P. L., & Lazarsfeld, P. F. (1950). Problems of survey analysis. In R. K. Merton & P. F. Lazarsfeld (Eds.), *Continuities in social research: Studies in the scope and method of "the American soldier"* (pp. 160-176). Glencoe, IL: Free Press.

Kennedy, P. (1992). *A guide to econometrics* (3rd ed.). Cambridge: MIT Press.

Kennedy, P. (1998). *A guide to econometrics* (4th ed.). Cambridge: MIT Press.

Kennedy, P. E. (2002). More on Venn diagrams for regression. *Journal of Statistics Education*, *10*(1) [Online]. Retrieved from http://www.amstat.org/publications/jse/v10n1/kennedy.html.

Kenny, D. A. (1979). *Correlation and causality.* New York: John Wiley.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis.* New York: Guilford.

Kenny, D. A., & La Voie, L. (1984). The social relations model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 18, pp. 142-182). Orlando, FL: Academic Press.

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Keren, G., & Lewis, C. (Eds.). (1993). *A handbook for data analysis in the behavioral sciences: Methodological issues.* Hillsdale, NJ: Lawrence Erlbaum.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research.* Ft. Worth, TX: Harcourt.

Kerr, A. W., Hall, H. K., & Kozub, A. (2002). Descriptive statistics. Chapter 2 in *Doing statistics with SPSS.* London: Sage.

Keselman, H. J., Lix, L. M., & Kowalchuk, R. K. (1998). Multiple comparison procedures for trimmed means. *Psychological Methods*, *3*, 123-141.

Kessler, R., & Greenberg, D. (1981). *Linear panel analysis.* New York: Academic Press.

Keyfitz, N. (1985). *Applied mathematical demography* (2nd ed.). New York: Wiley.

Keynes, J. M. (1936). *The general theory of employment, interest, and money.* New York: Harcourt Brace Jovanovich.

Khurana, B. (1995). *The older spouse caregiver: Paradox and pain of Alzheimer's disease.* Unpublished doctoral dissertation, Center for Psychological Studies, Albany, CA.

Kiel, L. D., & Elliott, E. (1996). *Chaos theory in the social sciences.* Ann Arbor: University of Michigan Press.

Kim, J. O., & Mueller, C. W. (1978). *Factor analysis: Statistical methods and practical issues.* Beverly Hills, CA: Sage.

Kinal, T., & Lahiri, K. (1983). Specification error analysis with stochastic regressors. *Econometrica*, *54*, 1209-1220.

Kincaid, H. (1994). Defending laws in the social sciences. In M. Martin & L. McIntyre (Eds.), *Readings in the philosophy of social science* (pp. 111-130). Cambridge: MIT Press.

Kincaid, H. (1996). *Philosophical foundations of the social sciences: Analyzing controversies in social research.* Cambridge, UK: Cambridge University Press.

King, G. (1988). Statistical models for political science event counts: Bias in conventional procedures and evidence for the exponential Poisson regression model. *American Journal of Political Science*, *32*, 838-863.

King, G. (1989). *Unifying political methodology: The likelihood theory of statistical inference.* New York: Cambridge University Press.

King, G. (1990). Stochastic variation: A comment on Lewis-Beck and Skalaban's "The R-Squared." *Political Analysis*, *2*, 185-200.

King, G. (1997). *A solution to the ecological inference problem.* Princeton, NJ: Princeton University Press.

King, G. (1998). *Unifying political methodo-logy: The likelihood theory of statistical inference.* Ann Arbor: University of Michigan Press.

King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, *9*, 137-163.

King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research.* Princeton, NJ: Princeton University Press.

King, J. A. (1998). Making sense of participatory evaluation practice. *New Directions for Evaluation*, *80*, 57-67.

Kirk, J., & Miller, M. L. (1986). *Reliability and validity in qualitative research.* Beverly Hills, CA: Sage.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Kirk, R. E. (1999). *Statistics: An introduction* (4th ed.). Orlando, FL: Harcourt Brace.

Kirkman, B. L., Rosen, B., Gibson, C. B., Tesluk, P. E., & McPherson, S. O. (2003). Five challenges to virtual team success: Lessons from Sabre, Inc. *Academy of Management Executive*, *16* (3), 67-79.

Kish, L. (1949). A procedure for objective respondent selection within a household. *Journal of the American Sociological Association*, *44*, 380-387.

Kish, L. (1962). Studies of interviewer variance for attitudinal variables. *Journal of the American Statistical Association*, *57*, 92-115.

Kish, L. (1965). *Survey sampling.* New York: Wiley.

Kish, L. (1987). *Statistical design for research.* New York: Wiley.

Kish, L. (1995). Methods for design effects. *Journal of Official Statistics*, *11*(1), 55-77.

Kitagawa, E. (1955). Components of a difference between two rates. *Journal of the American Statistical Association*, *50*, 1168-1194.

Kitcher, P., & Salmon, W. C. (Eds.). (1989). *Scientific explanation* (Minnesota Studies in the Philosophy of Science, Vol. 13). Minneapolis: University of Minnesota Press.

Kivlahan, D. R., Marlatt, G. A., Fromme, K., Coppel, D. B., & Williams, E. (1990). Secondary prevention with college drinkers: Evaluation of an alcohol skills training program. *Journal of Consulting and Clinical Psychology*, *58* (6), 805-810.

Kleiber, C., & Kotz, S. (2003). *Statistical size distributions in economics and actuarial sciences.* Hoboken, NJ: Wiley.

Kleinbaum, D. G., Kupper, L. L., & Morgensten, H. (1982). *Epidemiologic research: Principles and quantitative methods.* New York: Van Nostrand Reinhold.

Kleinman, A. L., Eisenberg, L., & Good, B. J. (1978). Culture, illness and care: Clinical lessons from anthropologic and cross-cultural research. *Annals of Internal Medicine*, *88*, 251-258.

Kleinman, S., & Copp, M. A. (1993). *Emotions and fieldwork.* Thousand Oaks, CA: Sage.

Kline, P. (1994). *An easy guide to factor analysis.* Thousand Oaks, CA: Sage.

Kline, R. B. (1998). *Principles and practice of structural equation modeling.* New York: Guilford.

Klir, G. J., & Folger, T. A. (1988). *Fuzzy sets, uncertainty, and information.* Englewood Cliffs, NJ: Prentice Hall.

Kmenta, J. (1997). *Elements of econometrics* (2nd ed.). Ann Arbor: University of Michigan Press.

Knoke, D., & Burke, P. J. (1980). *Log-linear models. Beverly Hills, CA: Sage.*

Knoke, D., Bohrnstedt, G. W., & Mee, A. P. (2002). *Statistics for social data analysis* (4th ed.). Itasca, IL: Peacock.

Knuth, D. E. (1997). *Fundamental algorithms: The art of computer programming* (Vol. 1, 3rd ed.). Reading, MA: Addison-Wesley.

Knuth, D. E. (1997). *The art of computer programming, volume 2: Seminumerical algorithms.* Reading, MA: Addison-Wesley.

Koch, W., Schulz, E. M., Wright, R., Smith, R. M., Lang, S.et al. (1996). What is a ratio scale? *Rasch Measurement Transactions*, *9*, 457.

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (Eds.). (2000). *To err is human: Building a safer health system.* Washington, DC: National Academy Press.

Kondo, D. K. (1990). *Crafting selves: Power, gender and discourses of identity in a Japanese workplace.* Chicago: University of Chicago Press.

Korczynski, M. (2000). The political economy of trust. *Journal of Management Studies*, *37*(1), 1-22.

Kornberg, A. (1997). *Basic research, the lifeline of medicine.* Retrieved from http://www. nobel. se/medicine/articles/research/.

Körner, S. (1955). *Kant.* Harmondsworth, UK: Penguin.

Kotz, S., & Johnson, N. (Eds.). (1983). *Encyclopedia of statistical sciences.* New York: Wiley.

Kotz, S., Johnson, N. L., & Read, C. B. (1982). Censoring. In S. Kotz, N. L. Johnson, & C. B. Read (Eds.), *Encyclopedia of statistical sciences* (Vol. 1, p. 396). New York: Wiley.

Koutsoyiannis, A. (1978). *Theory of econometrics: An introductory exposition of econometric methods* (2nd ed.).London: Macmillan.

Kozlowski, S. W., & Hattrup, K. (1992). A disagreement about within-group agreement: Disentangling issues of consistency versus consensus. *Journal of Applied Psychology*, *77*, 161-167.

Kracauer, S. (1952-1953). The challenge of qualitative content analysis. *Public Opinion Quarterly*, *16*, 631-642.

Krantz, J. H., & Dalal, R. (2000). Validity of Web-based psychological research. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 35-60). New York: Academic Press.

Krieger, N. (1994). Epidemiology and the web of causation: Has anyone seen the spider? *Social Science & Medicine*, *39*(7), 887-903.

Krieger, S. (1983). *The mirror's dance: Identity in a women's community.* Philadelphia: Temple University Press.

Krippendorf, K. (1980). *Content analysis: An introduction to its methodology.* Beverly Hills, CA: Sage.

Kristeva, J. (1991). *Strangers to ourselves.* New York: Columbia University Press.

Kroenke, D. M. (2001). *Database processing: Fundamentals, design and implementation.* Upper Saddle River, NJ: Prentice Hall.

Krosnick, J. A. (1999). Survey methodology. *Annual Review of Psychology*, *50*, 537-567.

Krueger, R. A. (1994). *Focus groups: A practical guide for applied research* (2nd ed.). Thousand Oaks, CA: Sage.

Krueger, R. A., & Casey, M. A. (2000). *Focus groups: A practical guide for applied research* (3rd ed.). Thousand Oaks, CA: Sage.

Kruskal, J. B. (1964). Multidimensional scaling by optimising goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*, 1-27, 115-129.

Kruskal, J. B. (1968). Transformation of data. In D. L. Sills (Ed.), *International encyclopedia of the social sciences* (Vol. 15, pp. 182-192). New York: Macmillan.

Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association*, *53*, 814-861.

Kuhn, M., & McPartland, T. S. (1954). An empirical investigation of self attitudes. *American Sociological Review*, *19*, 68-76.

Kuhn, T. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press. (Original work published 1962)

Kuran, T. (1995). The inevitability of future revolutionary surprises. *American Journal of Sociology*, *100*(6), 1528-1551.

Kuusela, H., Spence, M. T., & Kanto, A. J. (1998). Expertise effects on prechoice decision

processes and final outcomes: A protocol analysis. *European Journal of Marketing*, *32*,5-6, 559.

Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing.* Thousand Oaks, CA: Sage.

Kvale, S. (1999). The psychoanalytic interview as qualitative research. *Qualitative Inquiry*, *5*(1), 87-113.

Labov, W. (1982). Speech actions and reactions in personal narrative. In D. Tannen (Ed.), *Analyzing discourse: Text and talk.* Washington, DC: Georgetown University Press.

Lachenbruch, P. A. (1975). *Discriminant analysis.* New York: Hafner.

LaFreniere, P., & Charlesworth, W. R. (1983). Dominance, attention, and affiliation in a preschool group: A nine-month longitudinal study. *Ethology and Sociobiology*, *4*(2), 55-67.

Lagopoulos, A. Ph., & Boklund-Lagopoulou, K. (1992). *Meaning and geography: The social conception of the region in northern Greece.* Berlin: Mouton de Gruyter.

Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixture distribution. *Journal of the American Statistical Association*, *73*, 805-811.

Lakatos, I. (1978). *The methodology of scientific research programmes: Philosophical papers. Vol. 1.* Cambridge, UK: Cambridge University Press.

Lakatos, I., & Musgrave, A. (Eds.). (1970). *Criticism and the growth of know-ledge.* Cambridge, UK: Cambridge University Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.

Lambert, D. (1992). Zero-inflated Poisson regression with an application to defects in manufacturing. *Technometrics*, *34*, 1-14.

Langeheine, R., & van de Pol, F. (1990). A unifying framework for Markov modeling in discrete space and discrete time. *Sociological Methods & Research*, *18*, 416-441.

Langeheine, R., & Van de Pol, F. (1994). Discrete-time mixed Markov latent class models. In A. Dale & R. B. Davies (Eds.), *Analyzing social and political change: A casebook of methods* (pp. 171-197). London: Sage.

Langellier, K. M. (2001). Personal narrative. In M. Jolly (Ed.), *Encyclopedia of life writing: Autobiographical and biographical forms* (Vol. 2). London: Fitzroy Dearborn.

Langellier, K. M., & Peterson, E. E. (2003). *Performing marrative: The communicative practice of storytelling.* Philadelphia: Temple University Press.

Lapadat; J. C., & Lindsay, A. C. (1999). Transcription in research and practice: From standardization of technique to interpretive positionings. *Qualitative Inquiry*, *5*(1), 64-86.

Larsen, R. J., & Marx, M. L. (1990). *Statistics.* Englewood Cliffs, NJ: Prentice Hall.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior: The Hixon symposium* (pp. 112-136). New York: John Wiley.

Latham, G. P., Skarlicki, D., Irvine, D., & Siegel, J. P. (1993). The increasing importance of performance appraisals to employee effectiveness in organizational settings in North America. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology 1993* (Vol. 8, pp. 87-132). Chichester, UK: Wiley.

Lather, P. (1993). Fertile obsession: Validity after poststructuralism. *Sociological Quarterly*, *35*, 673-694.

Lather, P. (1999). To be of use: The work of reviewing. *Review of Educational Research*, *69*(1), 2-7.

Laudan, L. (1977). *Progress and its problems: Toward a theory of scientific growth.* Berkeley: University of California Press.

Lauder, M. (2003). Covert participant observation of a deviant community: Justifying the use of deception. *Journal of Contemporary Religion*, *18*(2), 185-196.

Laufer, R. S., & Wolfe, M. (1977). Privacy as a concept and a social issue: A multidimensional developmental theory. *Journal of Social Issues*, *33*, 44-87.

Laurie, H., Smith, R., & Scott, L. (1999). Strategies for reducing nonresponse in a longitudinal panel survey. *Journal of Official Statistics*, *15*(2), 269-282.

Lave, C., & March, J. G. (1978). *An introduction to*

models in the social sciences. New York: Harper & Row.

Lavrakas, P. J. (1987). *Telephone survey methods*. Newbury Park, CA: Sage.

Lavrakas, P. J. (1993). *Telephone survey methods: Sampling, selection, and supervision* (2nd ed.). Newbury Park, CA: Sage.

Lavrakas, P. J. (1998). Methods for sampling and interviewing in telephone surveys. In L. Bickman & D. J. Rog (Eds.), *Handbook of applied social research methods* (pp. 429-472). Thousand Oaks, CA: Sage.

Lawlis, G. F., & Lu, E. (1972). Judgment of counseling process: Reliability, agreement, and error. *Psychological Bulletin*, *78*, 17-20.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*, 563-575.

Lax, D. A. (1985). Robust estimators of scale: Finite-sample performance in long-tailed symmetric distributions. *Journal of the American Statistical Association*, *80*, 736-741.

Layder, D. (1993). *New strategies in social research*. Cambridge, UK: Polity.

Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis & the interpretation and mathematical foundation of latent structure analysis. In S. A. Stouffer. (Ed.), *Measurement and prediction* (pp. 362-472). Princeton, NJ: Princeton University Press.

Lazarsfeld, P. F. (1955). Interpretation of statistical relations as a research operation. In P. F. Lazarsfeld & M. Rosenberg (Eds.), *The language of social research* (pp. 111-125). Glencoe, IL: Free Press.

Lazarsfeld, P. F. (1958). Evidence and inference in social research. *Daedalus*, *87*, 120-121.

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.

Leach, C., Freshwater, K., Aldridge, J., & Sunderland, J. (2001). Analysis of repertory grids in clinical practice. *British Journal of Clinical Psychology*, *40*, 225-248.

Leamer, E. (1978). *Specification searches: Ad hoc inference with nonexperimental data*. New York: Wiley.

Leamer, E. E. (1978). *Specification searches: Ad hoc inference with nonexperimental data*. New York: John Wiley.

Leamer, E. E. (1983). Let's take the con out of econometrics. *American Economic Review*, *73*(1), 31-43.

Lebart, L., Morineau, A., & Warwick, K. M. (1984). *Multivariate descriptive statistical analysis*. New York: Wiley.

Lebreton, J.-D., Burnham, K. P., Clobert, J., & Anderson, D. R. (1992). Modeling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecological Monographs*, *62*, 67-118.

Lee, A. M. (1978). *Sociology for whom*. New York: Oxford University Press.

Lee, P. M. (1997). *Bayesian statistics: An introduction* (2nd ed.). London: Arnold.

Lee, R. (2000). *Unobtrusive methods in social research*. Buckingham, UK: Open University Press.

Lee, R. B. (1993). *The Dobe ! Kung*. New York: Holt, Rinehart and Winston.

Lee, R. M. (1993). *Doing research on sensitive topics*. London: Sage.

Lee, R. M. (1995). *Dangerous fieldwork*. London: Sage.

Lehtonen, M. (2000). *The cultural analysis of texts*. London: Sage.

Leisering, L., & Leibfried, S. (1999). *Time and poverty in Western welfare states*. New York: Cambridge University Press.

Lemaitre, G. (1992). *Dealing with the seam problem* (Survey of Labour and Income Dynamics Research Papers 92-05). Ottawa: Statistics Canada.

Lenski, G. (1966). *Power and privilege*. New York: McGraw-Hill.

Lepkowski, J. M., & Couper, M. P. (2002). Nonresponse in the second wave of longitudinal household surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 259-272). New York: John Wiley.

Leslie, P. H. (1945). On the use of matrices in certain population mathematics. *Biometrika*, *33*, 183-212.

Lett, J. (1987). The importance of the emic/etic

distinction. In *The human enterprise*: *A critical introduction to anthropological theory*. Boulder, CO: Westview.

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin (Ed.), *Contributions to probability and statistics* (pp. 278-292). Stanford, CA: Stanford University Press.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Cybernetics and Control Theory*, *10*, 707-710. (Original work published 1965)

Levin, A., Liukkonen, J., & Levine, D. W. (1996). Equivalent inference using transformations. *Communications in Statistics*, *Theory and Methods*, *25*(5), 1059-1072.

Levin, I. P. (1999). *Relating statistics and experimental design*: *An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-125). Thousand Oaks, CA: Sage.

Levin, J. R., & Subkoviak, M. J. (1977). Planning an experiment in the company of measurement error. *Applied Psychological Measurement*, *1*(3), 331-338.

Levinas, E. (1998). *On thinking-of-the-Other*: *Entre nous*. New York: Columbia University Press.

Levine, E. K. (1982). Old people are not all alike: Social class, ethnicity/race, and sex are bases for important differences. In J. E. Sieber (Ed.), *The ethics of social research*: *Surveys and experiments* (pp. 127-144). New York: Springer-Verlag.

Lévi-Strauss, C. (1963). *Structural anthropology*. New York: Basic Books.

Lévi-Strauss, C. (1969). *The elementary structures of kinship*. London: Eyre and Spottiswoode.

Levy, P. S., & Lemeshow, S. (1999). *Sampling of populations*: *Methods and applications*. New York: Wiley.

Lewin, K. (1943). Defining the "field at a given time." *Psychological Review*, *50*, 292-310.

Lewin, K. (1959). *A dynamic theory of personality*: *Selected papers* (D. K. Adams & K. E. Zener, Trans.). New York: McGraw-Hill. (Original work published 1935)

Lewin, K. (1997). Experiments in social space. In G. W. Lewin (Ed.), *Resolving social conflicts*: *Field theory in social science* (pp. 59-67). Washington, DC: American Psychological Association. (Originally published in 1939)

Lewis, D. K. (1973). *Counterfactuals*. Cambridge, MA: Harvard University Press.

Lewis-Beck, M. S. (1980). *Applied regression*: *An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, series 07-022). Beverly Hills, CA: Sage.

Lewis-Beck, M. S. (1995). *Data analysis*: *An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-103). Thousand Oaks, CA: Sage.

Lewis-Beck, M. S. (Ed.). (1993). *Regression analysis*. London: Sage/Toppan.

Lewis-Beck, M. S., & Skalaban, A. (1990). The *R*-squared: Some straight talk. *Political Analysis*, *2*, 153-171.

Li, H., Rosenthal, R., & Rubin, D. (1996). Reliability of measurement in psychology: From Spearman-Brown to maximal reliability. *Psychological Methods*, *1*(1), 98-107.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*, 13-22.

Liao, T. F. (1989). A flexible approach for the decomposition of rate differences. *Demography*, *26*, 717-726.

Liao, T. F. (1994). *Interpreting probability models*: *Logit*, *probit*, *and other generalized linear models*. Thousand Oaks, CA: Sage.

Liao, T. F. (2001). How responsive is U. S. population growth to immigration? A situational sensitivity analysis. *Mathematical Population Studies*, *9*, 217-229.

Liao, T. F. (2002). *Statistical group comparison*. New York: Wiley.

Lichfield, N. (1996). *Community impact evaluation*. London: University College Press.

Lieberson, S. (1987). *Making it count*: *The improvement of social research and theory*. Berkeley: University of California Press.

Light, R. J., & Pillemer, D. B. (1984). *Summing up*: *The science of reviewing research*. Cambridge, MA: Harvard University Press.

Likert, R. (1932). A technique for the measurement

of attitudes. *Archives of Psychology*, *140*, 44-53.

Lincoln, Y. S., & Guba, E. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

Lincoln, Y. S. & Guba, E. (2000). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 163-188). Thousand Oaks, CA: Sage.

Lind, J. (1753). *A treatise of the scurvy: Of three parts containing an inquiry into the nature, causes and cure of that disease*. Edinburgh, UK: Sands, Murray and Cochran.

Lindahl, L. (2002). *Do birth order and family size matter for intergenerational income mobility? Evidence from Sweden* (Working Paper No. 5). Stockholm: Swedish Institute for Social Research.

Lindell, M. K., & Brandt, C. J. (1999). Assessing interrater agreement on the job relevance of a test: A comparison of the CVI, T, $r_{WG(J)}$, and $r^*_{WG(J)}$ indexes. *Journal of Applied Psychology*, *84*, 640-647.

Lindell, M. K., Brandt, C. J., & Whitney, D. J. (1999). A revised index of interrater agreement for multi-item ratings of a single target. *Applied Psychological Measurement*, *23*, 127-135.

Lindesmith, A. (1968). *Addiction and opiates*. Chicago: Aldine.

Lindley, D. V. (2001). Thomas Bayes. In C. C. Heyde & E. Seneta (Eds.), *Statisticians of the centuries* (pp. 68-71). New York: Springer-Verlag.

Lindley, D. V., & Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society*, *34*, 1-41.

Lindquist, E. F. (1953). *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin.

Lindsay, B., Clogg, C. C., & Grego, J. (1991). Semiparametric estimation in the Rasch model and related models, including a simple latent class model for item analysis. *Journal of the American Statistical Association*, *86*, 96-107.

Lingis, A. (1994). *Abuses*. Berkeley: The University of California Press.

Lingis, A. (1994). *The community of those who have nothing in common*. Bloomington: Indiana University Press.

Lingis, A. (1998). *The imperative*. Bloomington: Indiana University Press.

Linn, R. L. (Ed.). (1989). *Educational measurement* (3rd ed.). New York: Macmillan.

Linstone, H. A. (1978). The Delphi technique In R. B. Fowles (Ed.), *Handbook of futures research*. Westport, CT: Greenwood.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger. R. D. (1996). *SAS system for mixed models*. Cary, NC: SAS Institute, Inc.

Little, D. (1991). *Varieties of social explanation: An introduction to the philosophy of social science*. Boulder, CO: Westview.

Little, D. (1998). *Microfoundations, method and causation: On the philosophy of the social sciences*. New Brunswick, NJ: Transaction Publishers.

Little, R. J. A. (1988). Missing data adjustments in large surveys. *Journal of Business and Economic Statistics*, *6*, 287-301.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: John Wiley.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York: John Wiley.

Litwin, M. S. (1995). *How to measure survey reliability and validity*. Thousand Oaks, CA: Sage.

Livingston, G. (1999). Beyond watching over established ways: A review as recasting the literature, recasting the lived. *Review of Educational Research*, *69*(1), 9-19.

Loader, C. (1999). *Local regression and likelihood*. New York: Springer.

Locke, K. D. (2000). *Using grounded theory in management research*. London: Sage.

Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3 (Suppl. 9), 635-694.

Loewenthal, L. (1989). Sociology of literature in retrospect. In P. Desan et al. (Eds.), *Literature and social practice*. Chicago and London: University of Chicago Press.

Lofland, J., & Lofland, L. H. (1995). *Analyzing social settings: A guide to qualitative observation*

and analysis. Belmont, CA: Wadsworth.

Loman, L. A., & Larkin, W. E. (1976). Rejection of the mentally ill: An experiment in labeling. *Sociological Quarterly*, *17*, 555-560.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.

Long, J. S., & Cheng, S. (2003). Regression models for categorical outcomes. In M. A. Hardy & A. Bryman (Eds.), *Handbook of data analysis*. London: Sage.

Long, J. S., & Ervin, L. H. (2000) Using heteroskedasticity consistent standard errors in the linear regression model. *American Statistician*, *54*, 217-224.

Longford, N. T. (1993). *Random coefficient models*. New York: Oxford University Press.

Looney, C. G. (1997). *Pattern recognition using neural networks*. Oxford, UK: Oxford University Press.

Lord, F. M. (1956). Sampling error due to choice of split in split-half reliability coefficients. *Journal of Experimental Education*, *24*, 245-249.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Lorencz, B. (1991). Becoming ordinary: Leaving the psychiatric hospital. In J. M. Morse & J. Johnson (Eds.), *The illness experience: Dimensions of suffering* (pp. 140-200). Newbury Park, CA: Sage.

Lorenz, E. N. (1963). Deterministic non-periodic flow. *Journal of Atmospheric Science*, *20*, 130-141.

Lorenz, K. (1981). *The foundations of ethology*. New York: Simon and Schuster.

Lotka, A. J. (1998). *Analytical theory of biological populations: Part II: Demographic analysis with particular application to human populations* (D. P. Smith & H. Rossert, Trans.). New York: Plenum. (Original work published 1939)

Luger, G. F. (2002). *Artificial intelligence: Structures and strategies for complex problem solving* (4th ed.). Reading, MA: Addison-Wesley.

Lukes, S. (1994). Methodological individualism reconsidered. In M. Martin & L. McIntyre (Eds.), *Readings in the philosophy of social science* (pp. 451-459). Cambridge: MIT Press.

Lundberg, G. (1939). *Foundations of sociology*. New York: Macmillan.

Lupia, A., McCubbins, M. D., & Popkin, S. L. (Eds.). (1998). *Elements of reason*. Cambridge, UK: Cambridge University Press.

Luttrell, W. (2003). *Pregnant bodies, fertile minds: Gender, race, and the schooling of pregnant teens*. New York: Routledge.

Lykken, D. T. (1957). A study of anxiety in the sociopathic personality. *Journal of Abnormal and Social Psychology*, *55*, 6-10.

Lykken, D. T. (1995). *The antisocial personalities*. Mahwah, NJ: Lawrence Erlbaum.

Lynch, M. (2000). Against reflexivity as an academic virtue and source of privileged knowledge. *Theory, Culture and Society*, *17*(3), 26-54.

Lyotard, J.-F. (1984). *The postmodern condition: A report on knowledge*. Minneapolis: University of Minnesota Press.

Macaulay, A. C., Paradis, G., Potvin, L., Cross, E. J., Saad-Haddad, C., McComber, A., et al. (1997). The Kahnawake Schools Diabetes Prevention Project: Intervention, evaluation, and baseline results of a diabetes primary prevention program with a native community in Canada. *Preventive Medicine*, *26*(6), 79-90.

MacDonald, I. L., & Zucchini, W. (1997). *Hidden Markov models and other types of models for discrete-valued time series*. London: Chapman & Hall.

MacDonald, R. R. (2002). The incompleteness of probability models and the resultant implications for theories of statistical inference. *Understanding Statistics*, *1*(3), 167-189.

MacDougall, D. (2001). Renewing ethnographic film: Is digital video changing the genre? *Anthropology Today*, *17*(3), 15-21.

Machlin, S. R., & Taylor, A. K. (2000). *Design, methods, and field results of the* 1996 *Medical Expenditure Panel Survey Medical Provider Component* (MEPS Methodology Report No. 9, AHRQ Pub. No. 00-0028). Rockville, MD: Agency for Healthcare Research and Quality.

Mackay, A. L. (1977). *Scientific quotations*. New York: Crane, Russak.

Mackie, J. L. (1974). *The cement of the universe: A*

study of causation. Oxford, UK: Oxford University Press.

Mackintosh, N. J. (1995). *Cyril Burt: Fraud or framed?* Oxford, UK: Oxford University Press.

Macy, M.W., & Willer, R. (2002). From factors to actors: Computational sociology and agent-based modeling. *Annual Review of Sociology*, *28*, 143-166.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge, UK: Cambridge University Press.

Maddala, G. S. (1992). *Introduction to econometrics*. Englewood Cliffs, NJ: Prentice Hall.

Madden, D. (2000). Towards a broader explanation of malefemale wage differences. *Applied Economics Letters*, *7*, 765-770.

Madjar, I. (2001). The lived experience of pain in the context of clinical practice. In S. K. Toombs (Ed.), *Handbook of phenomenology and medicine* (pp. 263-277). Boston: Kluwer Academic.

Magidson, J. (1993). The CHAID approach to segmentation modeling: CHi-squared Automatic Interaction Detection. In R. Bagozzi (Ed.), *Handbook of marketing research* (pp. 118-159). London: Blackwell.

Magidson, J., & Vermunt, J. K. (2001). Latent class factor and cluster models, bi-plots and related graphical displays. *Sociological Methodology*, *31*, 223-264.

Magnusson, D. (1967). *Test theory*. Reading, MA: Addison-Wesley.

Mahoney, J. (2000). Path dependence in historical sociology. *Theory and Society*, *29*, 507-548.

Maines, D. R. (1992). Theorizing movement in an urban transportation system by use of the constant comparative method in field research. *Social Science Journal*, *29*, 283-292.

Maines, D. R. (2001). *The faultline of consciousness: A view of interactionism in sociology*. New York: Aldine De Gruyter.

Makridakis, S., & Hibbon, M. (2000). The M3-competition: Results, conclusions, and implications. *International Journal of Forecasting*, *16*, 451-476.

Malinoswki, B. (1989). *A diary in the strict sense of the term*. Stanford, CA: Stanford University Press.

Malinowski, B. (1922). *Argonauts of the Western Pacific*. London: Routledge & Kegan Paul.

Manski, C. F. (1995). *Identification problems in the social sciences*. Cambridge, MA: Harvard University Press.

Manton, K. G., & XiLiang, G. (2003). *Variation in disability decline and Medicare expenditures*. Working Monograph, Center for Demographic Studies.

Manton, K. G., & Stallard, E. (1988). *Chronic disease modelling: Vol. 2. Mathematics in medicine*. New York: Oxford University Press.

Manton, K. G., Singer, B., & Woodbury, M. A. (1992). Some issues in the quantitative characterization of heterogeneous populations. In J. Trussell, R. Hankinson, & J. Tilton (Eds.), *Demographic application of event history analysis* (pp. 9-37). Oxford, UK: Clarendon.

Manton, K. G., Woodbury, M. A., & Tolley, H. D. (1994). *Statistical applications using fuzzy sets*. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley.

Marcus, G., & Fischer, M. (1986). *Anthropology as cultural critique*. Chicago: University of Chicago Press.

Marcuse, H. (1964). *One dimensional man*. Boston: Beacon.

Marcuse, H. (1968). *One dimensional man: The ideology of industrial society*. London: Sphere.

Mare, R. D. (1994). Discrete-time bivariate hazards with unobserved heterogeneity: A partially observed contingency table approach. In P. V. Marsden (Ed.), *Sociological methodology 1984* (pp. 341-383). Cambridge, MA: Blackwell.

Markham, A. N. (1998). *Life online: Researching real experience in virtual space*. Walnut Creek, CA: AltaMira.

Marquardt, D. W. (1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, *12*, 591-612.

Marquis, K. (1984). Record checks for sample surveys. In T. Jabine, E. Loftus, M. Straf et al. (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines*. Washington, DC: National Academy Press.

Marsh, C. (1982). *The survey method: The*

contribution of surveys to sociological explanation. London: Allen & Unwin.

Marsh, C. (1988). *Exploring data: An introduction to data analysis for social scientists.* Cambridge, UK: Polity.

Marsh, L. C., & Cormier, D. R. (2001). *Spline regression models.* Thousand Oaks, CA: Sage.

Marshall, P. A. (1992). Research ethics in applied anthropology. *IRB: A Review of Human Subjects Research*, *14*(6), 1-5.

Martens, H., & Naes, T. (1989). *Multivariate calibration.* London: Wiley.

Martin, J. A., Hamilton, B. E., Ventura, S. J., Menacker, F., & Park, M. M. (2002, February 12). *Births: Final data for 2000* (National Vital Statistics Report, Vol. 50, No. 5). Hyattsville, MD: National Center for Health Statistics.

Martin, K. (1998). *When a baby dies of SIDS: The parents' grief and searching for a reason.* Edmonton, Canada: Qual Institute Press.

Mason, J. (1996). *Qualitative researching.* London: Sage.

Mason, J. (2002). *Qualitative researching* (2nd ed.). London: Sage.

Mason, K. O., Mason, W. M., Winsborough, H. H., & Poole, W. K. (1973). Some methodological issues in the cohort analysis of archival data. *American Sociological Review*, *38*, 242-258.

Mathiowetz, N., & McGonagle, K. (2000). An assessment of the current state of dependent interviewing in household surveys. *Journal of Official Statistics*, *16*(4), 401-418.

Matza, D. (1969). *Becoming deviant.* Englewood Cliffs, NJ: Prentice-Hall.

Maxim, P. S. (1999). *Quantitative research methods in the social sciences.* New York: Oxford University Press.

Maxwell, J. A. (1996). *Qualitative research design.* Thousand Oaks, CA: Sage.

Maxwell, S. E., & Delaney, H. D. (1990). *Design experiments and analyzing data: A model comparison perspective.* Belmont, CA: Wadsworth.

May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature*, *26*, 459-467.

Maynard, D. W. (1984). *Inside plea bargaining: The language of negotiation.* New York: Plenum.

McAdams, D. P., & Constantian, C. A. (1983). Intimacy and affiliation motives in daily living: An experience sampling analysis. *Journal of Personality and Social Psychology*, *45*, 851-861.

McCleary, R., & Hay, R. A., Jr. (1980). *Applied time series analysis for the social sciences.* London: Sage.

McCloskey, D. N. (1986). *The rhetoric of economics.* Madison: University of Wisconsin Press.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B*, *42*, 109-142.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

McCulloch, C. E., & Searle, S. R. (2001). *Generalized, linear and mixed models.* New York: John Wiley.

McCullough, B. D., & Vinod, H. D. (1997). The numerical reliability of econometric software. *Journal of Economic Literature*, *37*(2), 633-665.

McDonald, J. F., & Moffitt, R. A. (1980). The uses of Tobit analysis. *Review of Economics and Statistics*, *62*, 318-321.

McDowall, D., McCleary, R., Meidinger, E. E., & Hay, R. A., Jr. (1980). *Interrupted time series analysis.* Beverly Hills, CA: Sage.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers of econometrics* (pp. 105-142). New York: Academic Press.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*, 30-46.

McGrew, W. (1972). *An ethological study of children's behavior.* New York: Academic Press.

McGuigan, J. (1992). *Cultural populism.* London: Routledge.

McGuinness, M., & Sapsford, R. (2002). *Middlesbrough Citizens' Advice Bureau: Report for 2001.* Middlesbrough, UK: University of Teesside, School of Social Sciences.

McIntosh, A. R., Bookstein, F. L., Haxby, J. V., & Grady, C. L. (1996). Spatial pattern analysis of functional brain images using partial least squares.

Neuroimage, *3*, 143-157.

McIver, J. P., & Carmines, E. G. (1980). *Unidimensional scaling.* Beverly Hills, CA: Sage.

McKean, K. (1987, January). The orderly pursuit of pure disorder. *Discover*, pp. 72-81.

McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, *4*, 103-120.

McLachlan, G. J., & Basford, K. E. (1988). *Mixture models: Inference and application to clustering.* New York: Marcel Dekker.

McSweeny, A. J. (1978). Effects of response cost on the behavior of a million persons: Charging for directory assistance in Cincinnati. *Journal of Applied Behavior Analysis*, *11*, 47-51.

Meacham, S. J. (1998). Threads of a new language: A response to Eisenhart's "On the subject of interpretive reviews." *Review of Educational Research*, *68*(4), 401-407.

Mead, M. (1928). *Coming of age in Samoa.* New York: William Morrow.

Meadows, L. M., & Dodendorf, D. M. (1999). Data management and interpretation: Using computers to assist. In B. F. Crabtree & W. L. Miller (Eds.), *Doing qualitative research* (pp. 195-218). Thousand Oaks, CA: Sage.

Melton, G. B., Levine, R. J., Koocher, G. P., Rosenthal, R., & Thompson, W. C. (1988). Community consultation in socially sensitive research: Lessons from clinical trials on treatments for AIDS. *American Psychologist*, *43*, 573-581.

Menard, S. (1995). *Applied logistic regression analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-106). Thousand Oaks, CA: Sage.

Menard, S. (2002). *Applied logistic regression analysis* (2nd ed., Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-106). Thousand Oaks, CA: Sage.

Menard, S. (2002). *Longitudinal research* (2nd ed.). Thousand Oaks, CA: Sage.

Menard, S. (in press). *Logistic regression.* Thousand Oaks, CA: Sage.

Menard, S., & Elliott, D. S. (1990). Longitudinal and crosssectional data collection and analysis in the study of crime and delinquency. *Justice Quarterly*, *7*, 11-55.

Menchú, R. (1984). *I, Rigoberta Menchú: An Indian woman in Guatemala.* (Ed. and intro. Elisabeth Burgos-Debray, Trans. Ann Wright.) London: Verso.

Mendel, G. (1866). Versuche über Pflanzen-Hybriden [Experiments in plant hybridization]. *In Verhandlungen des naturforschenden Vereins* [Proceedings of the Natural History Society]. Available in both the original German and the English translation at www.mendelweb.org.

Merkle, D. M., & Edelman, M. (2002). Nonresponse in exit polls: A comprehensive analysis. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 243-258). New York: Wiley.

Merleau-Ponty, M. (1962). *Phenomenology of perception.* London: Routledge Kegan Paul.

Merton, R. K. (1949). *Social theory and social structure.* New York: Free Press.

Merton, R. K. (1968). *Social theory and social structure* (3rd ed.). New York: Free Press.

Merton, R. K. (1987). The focused interview and focus groups: Continuities and discontinuities. *Public Opinion Quarterly*, *51*, 550-566.

Merton, R. K., & Kendall, P. L. (1946). The focused interview. *American Journal of Sociology*, *51*(6), 514-557.

Merton, R. K., Fiske, M., & Kendall, P. L. (1990). *The focused interview: A manual of problems and procedures* (2nd ed.). New York: Free Press.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, *18*(2), 5-11.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: American Council on Education.

Messick, S. (1992). Validity of test interpretation and use. In M. C. Alkin (Ed.), *Encyclopedia of educational research* (6th ed., p. 1487-1495). New York: Macmillan.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry

into score meaning. *American Psychologist*, *50*(9), 741-749.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculation by fast computing machines. *Journal of Chemical Physics*, *21*, 1087-1092.

Meulman, J. J. (1998). Book review of W. J. Krzanowski, & F. H. C. Marriott, *Multivariate Analysis. Part I. Distributions, Ordinations, and Inference*, London: Edward Arnold, 1994. *Journal of Classification*, *15*, 287-293.

Meulman, J. J., & Heiser, W. J. (2000). *Categories.* Chicago: SPSS.

Michel, Y., & Haight, B. K. (1996). Using the Solomon four design. *Nursing Research*, *45*(6), 367-369.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, *67*, 371-378.

Milgram, S. (1974). *Obedience to authority.* London: Tavistock.

Milgram, S. (1974). *Obedience to authority: An experimental view.* New York: Harper & Row.

Mill, J. S. (1947). *A system of logic.* London: Longman, Green. (Originally published in 1879)

Mill, J. S. (1956). *A system of logic, ratiocinative and inductive.* London: Longmans, Green & Company. (Original work published 1843)

Mill, J. S. (1973). *A system of logic.* Toronto: University of Toronto Press. (Original work published 1865)

Mill, J. S. (1974). *Collected works of John Stuart Mill: Vol. 8. A system of logic* (Books 4-6). Toronto: University of Toronto Press.

Millar, A., Simeone, R. S., & Carnevale, J. T. (2001). Logic models: A system tool for performance management. *Evaluation and Program Planning*, *24*, 73-81.

Miller, A. G. (1986). *The obedience experiments: A case study of controversy in social science.* New York: Praeger.

Miller, A. G. (1995). Constructions of the obedience experiments: A focus upon domains of relevance. *Journal of Social Issues*, *51*, 33-53.

Miller, A. J. (1990). *Subset selection in regression.* New York: Chapman Hall.

Miller, D. (1998). Writing and retelling multiple ethnographic tales of a soup kitchen for the homeless. *Qualitative Inquiry*, *4*, 469-492.

Miller, L., Rustin, M., Rustin, M., & Shuttleworth, J. (1989). *Closely observed infants.* London: Duckworth.

Miller, R. G. (1981). *Simultaneous statistical inference* (2nd ed.). New York: Springer-Verlag.

Miller, R. W. (1987). *Fact and method: Explanation, confirmation and reality in the natural and the social sciences.* Princeton, NJ: Princeton University Press.

Miller, W. L., & Crabtree, B. F. (1999). Clinical research: A multimethod typology and qualitative roadmap. In B. F. Crabtree & W. L. Miller (Eds.), *Doing qualitative research* (2nd ed., pp. 3-32). Thousand Oaks, CA: Sage.

Millett, K. (1970). *Sexual politics.* Garden City, NY: Doubleday.

Milligan, G. W. (1979). Ultrametric hierarchical clustering algorithms. *Psychometrika*, *44*(3), 343-346.

Milliken, G. A., & Johnson, D. A. (1984). *Analysis of messy data: Vol. 1. Designed experiments.* New York: Van Nostrand Reinhold.

Milliken, G. A., & Johnson, D. E. (1992). *Analysis of messy data: Vol. 1. Designed experiments.* London: Chapman & Hall.

Mills, C. W. (1959). *The sociological imagination.* New York: Oxford University Press.

Mills, T. C. (1990). *Time series techniques for economists.* Cambridge, UK: Cambridge University Press.

Millward, L. J. (2000). Focus groups. In G. M. Breakwell, C. Fife-Schaw, & S. Hammond (Eds.), *Research methods in psychology* (2nd ed., pp. 303-324). London: Sage.

Milner, A. (1996). *Literature, culture and society.* London: UCL Press.

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy*, *66*, 281-302.

Mincer, J. (1974). *Schooling, experience and*

earnings. New York: Columbia University Press.

Mishler, E. G. (1986). *Research interviewing: Context and narrative*. Cambridge, MA: Harvard University Press.

Mishler, E. G. (1991). Representing discourse: The rhetoric of transcription. *Journal of Narrative and Life History*, *1*(4), 255-280.

Mishler, E.G. (1995). Models of narrative analysis: A typology. *Journal of Narrative and Life History*, *5*(2), 87-123.

Mitchell, J. (1983). Case and situation analysis. *Sociological Review*, *31*(2), 186-211.

Mitchell, J. (1986). Measurement scales and statistics: A clash of paradigms. *Psychological Bulletin*, *100*, 398-407.

Mittelhammer, R. C., Judge, G. G, & Miller, D. J. (2000). *Econometric foundations*. NewYork: Cambridge University Press.

Mohr, D. C., Likosky, W., Bertagnolli, A., Goodkin, D. E., Van der Wende, J., Dwyer, P., et al. (2000). Telephone administered cognitive-behavioral therapy for the treatment of depressive symptoms in multiple sclerosis. *Journal of Consulting and Clinical Psychology*, *68*(2), 356-361.

Mohr, L. B. (1990). *Understanding significance testing* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-073). Newbury Park, CA: Sage.

Monge, P., & Contractor, N. (2003). *Theories of communication networks*. New York: Oxford University Press.

Monk-Turner, E., & Turner, C. G. (2001). Sex differentials in the South Korean labour market. *Feminist Economics*, *7*(1), 63-78.

Mood, A. M., & Graybill, F. A. (1963). *Introduction to the theory of statistics* (2nd ed.). New York: McGraw-Hill.

Mooney, C. Z. (1997). *Monte Carlo simulation*. Thousand Oaks, CA: Sage.

Mooney, C. Z., & Duval, R. D. (1993). *Bootstrapping: A nonparametric approach to statistical inference*. Thousand Oaks, CA: Sage.

Moore, D. S. (1997). *Statistics: Concepts and controversies* (4th ed.). New York: W. H. Freeman.

Moore, D. S., & McCabe, G. P. (1998). *Introduction to the practice of statistics*. New York: W. H. Freeman.

Moore, D. W. (1999, June/July). Daily tracking polls: Too much "noise" or revealed insights? *Public Perspective*, pp. 27-31.

Moorman, R. H., & Podsakoff, P. M. (1992). A meta-analytic review and empirical test of the potential confounding effects of social desirability response sets in organizational behaviour research. *Journal of Occupational and Organizational Psychology*, *65*, 131-149.

Moreno, L. (1994). Frailty selection in bivariate survival models: A cautionary note. *Mathematical Population Studies*, *4*, 225-233.

Morgan, D. L. (1988). *Focus groups as qualitative research*. Newbury Park, CA: Sage.

Morgan, D. L. (1993). *Successful focus groups: Advancing the state of the art*. London: Sage.

Morgan, D. L. (1998). Practical strategies for combining qualitative and quantitative methods: Applications for health research. *Qualitative Health Research*, *8*, 362-376.

Morgan, D. L., & Krueger R. A. (Eds.). (1998). *Focus group kit*. Thousand Oaks, CA: Sage.

Morley, D. (1980). *The "Nationwide" audience*. London: British Film Institute.

Morley, D. (1986). *Family television—Cultural power and domestic leisure*. London: British Film Institute.

Morris, M., & Western, B. (1999). Inequality in earnings at the close of the 20th century. *Annual Review of Sociology*, *25*, 623-657.

Morrison, D. E., & Henkel, R. (Eds.). (1970). *The significance test controversy*. Chicago: Aldine.

Morrow, R. A. (with Brown, D. D.). (1994). *Critical theory and methodology*. London: Sage.

Morse, J. M. (1989). Strategies for sampling. In J. Morse (Ed.), *Qualitative nursing research: A contemporary dialogue* (pp. 117-131). Rockville, MD: Aspen.

Morse, J. M. (1997). Considering theory derived from qualitative research. In J. Morse (Ed.), *Completing a qualitative project: Details and dialogue* (pp. 2163-2188). Thousand Oaks, CA: Sage.

Morse, J. M., & Richards, L. (2002). *Read me first*

for a user's guide to qualitative methods. Thousand Oaks, CA: Sage.

Morse, J. M., Swanson, J. M., & Kuzel, A. J. (Eds.). (2001). *The nature of qualitative evidence*. Thousand Oaks, CA: Sage.

Morton, Rebecca B. (2000). *Methods and models: A guide to the empirical analysis of formal models in political science*. New York: Cambridge University Press.

Moser, C. A., & Kalton, G. (1971). *Survey methods in social investigation* (2nd ed.). Aldershot, UK: Gower.

Mosteller, F., & Tukey, J. W. (1977). *Data analysis and regression: A second course in statistics*. Reading, MA: Addison-Wesley.

Mounce, H. O. (1997). *The two pragmatisms: From Peirce to Rorty*. London: Routledge and Kegan Paul.

Moustakas, C. (1990). *Heuristic research: Design, methodology, and applications*. Newbury Park, CA: Sage.

Moustakas, C. (1994). *Phenomenological research methods*. Thousand Oaks, CA: Sage.

Moyé, L. A. (2000). *Statistical reasoning in medicine*. New York Springer-Verlag.

Mruck, K., Corti, L., Kluge, S. & Opitz, D. (Eds.). (2000, December). Text archive: Re-analysis. *Forum Qualitative Sozialforschung* [Forum: Qualitative Social Research] [Online Journal], *1* (3). Available: http://qualitativeresearch.net/fqs/fqs-eng.htm.

Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement*, *56*, 63-75.

Mueller, D. J. (1986). *Measuring social attitudes: A handbook for researchers and practitioners*. New York: Teachers College, Columbia University.

Mueller, R. O. (1996). *Basic principles of structural equation modeling: An introduction to LISREL and EQS*. New York: Springer.

Mulaik, S. A. (1972). *The foundations of factor analysis*. New York: McGraw-Hill.

Mulaik, S. A. (1987). A brief history of the philosophical foundations of exploratory factor analysis. *Multivariate Behavioral Research*, *22*, 267-305.

Murphy, M. (1990). Minimising attrition in longitudinal studies: Means or end? In D. Magnusson & L. R. Bergman (Eds.), *Data quality in longitudinal research* (pp. 148-156). Cambridge, UK: Cambridge University Press.

Murphy, M. K., Black, N. A., Lamping, D. L., McKee, C. M., Sanderson, C. F. B., Ashkam, J., & Marteau, T. (1998). Consensus development methods, and their use in clinical guideline development [Entire issue]. *Health Technology Assessment*, *2*(3).

Musch, J., & Reips, U. (2000). A brief history of Web experimenting. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 61-88). San Diego, CA: Academic Press.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*, 115-132.

Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, *29* (1), 81-117.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, *51*, 431-462.

Myers, J. L., & Well, A. D. (1995). *Research design and statistical analysis*. Hillsdale, NJ: Lawrence Erlbaum.

Naes, T., & Risvik, E. (Eds.). (1996). *Multivariate analysis of data in sensory science*. New York: Elsevier.

Nagel, E. (1979). *The structure of science: Problems in the logic of scientific explanation*. Indianapolis, IN: Hackett.

Nagler, J. (1995). Coding style and good computing practices. *Political Methodologist*, *6*, 2-8.

Namboodiri, K. (1984). *Matrix algebra: An introduction*. Beverly Hills, CA: Sage.

Narayan, K. (1993). How native is the "native" anthropologist? *American Anthropologist*, *95*, 671-686.

National Center for Health Statistics. (2003). *United States life tables, 2000*. Retrieved from www.cdc.gov/nchs/data/ lt2000.pdf.

National Commission for the Protection of Human

Subjects of Biomedical and Behavioral Research. (1979). *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research* (DHEW Publication No. (OS) 78-0012). Washington, DC: Government Printing Office.

National Institutes of Health. (2003, July 22). *Certificates of confidentiality.* Retrieved from http://grants1.nih.gov/grants/ policy/coc/index.htm.

National Research Council. (2001). *The 2000 census: Interim assessment.* Washington, DC: National Academy Press.

Naughton, J. (1984). *Soft systems analysis: An introductory guide.* Milton Keynes, UK: Open University Press.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology.* San Francisco: W. H. Freeman.

Nelder, J. A., & Lane, P. W. (1995). The computer analysis of factorial experiments: In memoriam—Frank Yates. *The American Statistician*, *49*, 382-385.

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, *Series A*, *135*, 370-384.

Nelson, J. S., Megill, A., & McCloskey, D. N. (Eds.). (1987). *The rhetoric of the human sciences.* Madison: University of Wisconsin Press.

Nesbary, D. K. (2000). *Survey research and the World Wide Web.* Boston, MA: Allyn and Bacon.

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear regression models* (3rd ed.). Chicago: Irwin.

Neter, J., Wasseman, W., & Kutner, M. H. (1989). *Applied linear regression models* (2nd ed.). Boston: Irwin.

Neter, J., Wasserman, W., & Kutner, M. (1990). *Applied linear statistical models* (3rd ed.). Homewood, IL: Irwin.

Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement*, *22*, 287-293.

New, C. (1996). *Agency, health and social survival.* London: Taylor & Francis.

Newbold, P., & Bos, T. (1985). *Stochastic parameter regression models* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-051). Beverly Hills, CA: Sage.

Newbold, P., & Bos, T. (1994). *Introductory business forecasting* (2nd ed.). Cincinnati, OH: South-Western.

Newby, H. (1977). In the field: Reflections on a study of Suffolk farm workers. In C. Bell & H. Newby (Eds.), *Doing sociological research* (pp. 108-129). London: Allen and Unwin.

Newell, A., & Simon, H. A. (1961). Computer simulation of human thinking. *Science*, *134*, 2011-2017.

Newell, A., & Simon, H. A. (1971). *Human problem solving.* Englewood Cliffs, NJ: Prentice Hall.

Newell, R. W. (1986). *Objectivity, empiricism and truth.* London: Routledge and Kegan Paul.

Newey, W. K., & West, K. D. (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*, 703-708.

Newton, I. (1952). *Mathematical principles of natural philosophy.* Chicago: Britannica. (Originally published in 1686)

Newton, R. R., & Rudestam, K. E. (1999). *Your statistical consultant: Answers to your data analysis questions.* Thousand Oaks, CA: Sage.

NHS Centre for Reviews and Dissemination. (2001). *Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews* (Report number 4, 2nd ed.). York, UK: CRD.

Nicholson, L. (Ed.). (1990). *Feminism/postmodernism.* New York: Routledge Kegan Paul.

Nida, E. A. (1975). *Componential analysis of meaning: An introduction to semantic structures.* The Hague, The Netherlands: Mouton.

Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications.* Toronto: University of Toronto Press.

Nishisato, S. (1994). *Elements of dual scaling: An introduction to practical data analysis.* Hillsdale, N. J.: Lawrence Erlbaum.

Nishisato, S. (1996). Gleaning in the field of dual scaling. *Psychometrika*, *61*, 559-599.

Nishisato, S., & Clavel, J. G. (2003). A note on betweenset distances in dual scaling and

correspondence analysis. *Behaviormetrika*, *30*, 87-98.

NIST/SEMATECH. (2003). *E-handbook of statistical methods*. Retrieved from http://www.itl.nist.gov/div898/handbook/eda.

Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography: Synthesizing qualitative studies*. Newbury Park, CA: Sage.

Nocedal, J., & Wright, S. (1999). *Numerical optimization*. New York: Springer-Verlag.

Noldus, L. P. J. J., Trienes, R. J. H., Hendriksen, A. H. M., Jansen, H., & Jansen, R. G. (2000). The Observer Video-Pro: New software for the collection, management, and presentation of time-structured data from videotapes and digital media files. *Behavior Research Methods, Instruments & Computers*, *32*, 197-206.

Norpoth, H. (1995). Is Clinton doomed? An early forecast for 1996. *Political Science & Politics*, *27*, 201-206.

Norton, H. W. (1939). The 7 × 7 squares. *Annals of Eugenics*, *9*, 269-307.

Norusis, M. (1990). *SPSS/PC + Statistics 4.0*. Chicago: SPSS.

Norusis, M. J. (1990). *SPSS base system user's guide*. Chicago: SPSS.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). E-research: Ethics, security, design, and control in psychological research on the Internet. *Journal of Social Issues*, *58*, 161-176.

Nowak, A., & Latané, B. (1994). Simulating the emergence of social order from individual behaviour. In N. Gilbert & J. Doran (Eds.), *Simulating societies: The computer simulation of social phenomena* (pp. 63-84). London: UCL.

Noymer, A. (2001). The transmission and persistence of "urban legends": Sociological application of age-structured epidemic models. *Journal of Mathematical Sociology*, *25*, 299-323.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Nutley, S., & Webb, J. (2000). Evidence and the policy process. In H. Davies, S. Nutley, & P. Smith (Eds.), *What works? Evidence-based policy and practice in public services* (pp. 13-41). Bristol, UK: Policy Press.

O'Brien, R. M. (2000). Age-period-cohort-characteristic models. *Social Science Research*, *29*, 123-139.

O'Hear, A. (1989). *An introduction to the philosophy of science*. Oxford, UK: Clarendon.

O'Hear, A. (Ed.). (1996). *Verstehen and humane understanding*. Supplement to *Philosophy*, Royal Institute of Philosophy Supplement 41. Cambridge, UK: Cambridge University Press.

O'Rourke, D., & Blair, J. (1983). Improving random respondent selection in telephone surveys. *Journal of Marketing Research*, *20*, 428-432.

Oakley, A. (2000). *Experiments in knowing: Gender and method in the social sciences*. Cambridge, UK: Polity.

Ohnuki-Tierney, E. (1984). "Native" anthropologists. *American Ethnologist*, *11*, 584-586.

Oksenberg, L., Cannell, C., & Kalton, G. (1991). New strategies for pretesting survey questions. *Journal of Official Statistics*, *7*(3), 349-365.

Olmsted, P., & Weikart, D. P. (Eds.). (1994). *Families speak: Early childhood care and education in 11 countries*. Ypsilanti, MI: High/Scope Press.

Olsen W. K. (1993) Random samples and repeat surveys in South India. In S. Devereux & J. Hoddinott (Eds.) *Fieldwork in Developing Countries*, (pp. 57-72). Boulder, CO: Lynne Rienner.

Olsen, W. K. (1999). *Path analysis for the study of farming and micro-enterprise* (Bradford Development Paper No. 3). Bradford, UK: University of Bradford, Development and Project Planning Centre.

Olsen, W. K., Warde, A., & Martens, L. (2000). Social differentiation and the market for eating out in the UK. *International Journal of Hospitality Management*, *19*(2), 173-190.

Oneal, J. R., & Russett, B. (1997). The classical liberals were right: Democracy, interdependence, and conflict, 1950-1985. *International Studies Quarterly*, *41*, 267-294.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference

to demand characteristics and their implications. *American Psychologist*, *17*, 776-783.

Orne, M. T. (1969). Demand characteristics and the concept of quasi-controls. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 143-179). New York: Academic Press.

Orr, L. L. (1999). *Social experiments: Evaluating public programs with experimental methods.* Thousand Oaks, CA: Sage.

Orwin, R. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*, 157-159.

Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods*, *5*, 343-355.

Osgood, C. E. (1952). The nature of and measurement of meaning. *Psychological Bulletin*, *49*, 197-237.

Osgood, C. E., & Tzeng, O. C. S. (Eds.). (1990). *Language, meaning, and culture: The selected papers of C. E. Osgood.* New York: Praeger.

Osgood, C. E., Suci, C. J., & Tannenbaum, P. H. (1957). *The measurement of meaning.* Urbana: University of Illinois Press.

Ostrom, C. W., Jr. (1990). *Time series analysis.* Thousand Oaks, CA: Sage.

Otis, D. L., Burnham, K. P., White, G. C., & Anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs*, *62*, 1-135.

Outhwaite, W. (1975). *Understanding social life: The method called verstehen.* London: Allen & Unwin.

Outhwaite, W. (1987). *New philosophies of social science: Realism, hermeneutics and critical theory.* London: Macmillan.

Paccagnella, L. (1997). Getting the seats of your pants dirty: Strategies for ethnographic research on virtual communities. *Journal of Computer Mediated Communication*, *3* (*1*). Retrieved from http://www. ascusc. org/jcmc/vol3/issue1/paccagnella. html.

Paechter, C. (1996). Power, knowledge and the confessional in qualitative research. *Discourse: Studies in the Politics of Education*, *17*(1), 75-84.

Page, B. I., & Shapiro, R. Y. (1992). *The rational public.* Chicago: University of Chicago Press.

Page, S. (2000). The lost art of unobtrusive measures. *Journal of Applied Social Psychology*, *30* (10), 2126-2128.

Pagès, J., & Tenenhaus, M. (2001). Multiple factor analysis combined with PLS path modeling: Application to the analysis of relationships between physicochemical variables, sensory profiles and hedonic judgments. *Chemometrics and Intelligent Laboratory Systems*, *58*, 261-273.

Palmer, R. E. (1969). *Hermeneutics: Interpretation theory of Schleiermacher, Dilthey, Heidegger, and Gadamer.* Evanston, IL: Northwestern University Press.

Pampel, F. C. (2000). *Logistic regression: A primer.* Thousand Oaks, CA: Sage.

Pankratz, L. (1979). Symptom validity testing and symptom retraining: Procedures for the assessment and treatment of functional sensory deficits. *Journal of Consulting and Clinical Psychology*, *47*, 409-410.

Papineau, D. (1993). *Philosophical naturalism.* Oxford, UK: Basil Blackwell.

Park, A., & Jowell, R. (1997). *Consistencies and differences in a cross-national survey.* London: SCPR.

Parker, I. (1988). Deconstructing accounts. In C. Antaki (Ed.), *Analysing everyday explanation: A casebook of methods* (pp. 184-198). London: Sage.

Parkin, F. (1973). *Class inequality and political order.* London: Paladin.

Parkinson, B., & Manstead, A. S. R. (1993). Making sense of emotion in stories and social life. *Cognition and Emotion*, *7*, 295-323.

Parsons, T. (1968). Émile Durkheim. In D. L. Sills (Ed.), *International encyclopedia of the social sciences*, *Vol. 4.* New York: Macmillan.

Parsons, T. (1968). *The structure of social action* (3rd ed.). New York: Free Press.

Passmore, J. (1967). Logical positivism. In P. Edwards (Ed.), *The encyclopedia of philosophy* (Vol. 5, pp. 52-57). New York: Macmillan.

Paterson, B. L., Thorne, S. E., Canam, C., & Jillings, C. (2001). *Meta-study of qualitative health research: A practical guide to meta-analysis and meta-synthesis.* Thousand Oaks, CA: Sage.

Patrick, C. J. (1994). Emotion and psychopathy:

Startling new insights. *Psychophysiology*, *31*, 415-428.

Pattison, P. E. (1993). *Algebraic models for social networks.* New York: Cambridge University Press.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: Sage.

Patton, M. Q. (1999). *Grand Canyon celebration: A father-son journey of discovery.* Amherst, NY: Prometheus.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*, 598-609.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation.* London: Sage.

Pearl, J. (2000). *Causality: Models, reasoning, and inference.* Cambridge, UK: Cambridge University Press.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, *6*(2), 559-572.

Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *British Medical Journal*, *3*, 1243-1246.

Peat, J., Mellis, C., Williams, K., & Xuan W. (2002). *Health science research: A handbook of quantitative methods.* London: Sage.

Pedhazer, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach.* Hillsdale, NJ: Lawrence Erlbaum.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd ed.). New York: Holt, Rinehart and Winston.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research* (3rd ed.). New York: Wadsworth.

Peirce, C. S. (1931-1958). *Collected papers of Charles Sanders Peirce.* Cambridge, MA: Harvard University Press.

Pelusi, J. (1997). The lived experience of surviving breast cancer. *Oncology Nursing Forum*, *24*(8), 1343-1353.

Perks, R., & Thomson, A. (Eds.). (1998). *The oral history reader.* London: Routledge.

Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research*, *21*, 381-391.

Pettofrezzo, A. J. (1978). *Matrices and transformations.* New York: Dover.

Pfungst, O. (1965). *Clever Hans: The horse of Mr. von Osten.* New York: Holt, Rinehart, & Winston. (Original work published 1911)

Phillips, D. (1976). *Holistic thought in social science.* Stanford, CA: Stanford University Press.

Phillips, N., & Brown, J. (1993). Analyzing communication in and around organizations: A critical hermeneutic approach. *Academy of Management Journal*, *36*(6), 1547-1576.

Phillips, P. C. B. (1991). Optimal inference in cointegrated systems. *Econometrica*, *59*, 283-306.

Piaget, J. (1929). *The child's conception of the world.* New York: Harcourt, Brace Jovanovich.

Piaget, J. (1955). *The construction of reality in the child.* New York: Routledge Kegan Paul. (Original work published 1937)

Pickering, M. (2001). *Stereotyping: The politics of representation.* Basingstoke, UK: Palgrave.

Pierce, J. L., Gardner, D. G., Cummings, L. L., & Dunham, R. B. (1989). Organization-based self-esteem: Construct definition, measurement, and validation. *Academy of Management Journal*, *32*, 622-648.

Pierson, P. (2000). Increasing returns, path dependence, and the study of politics. *American Political Science Review*, *94*, 251-267.

Pike, K. L. (1967). *Language in relation to a unified theory of the structure of human behavior* (2nd ed.). The Hague: Mouton. (Original work published 1954)

Pindyck, R. S., & Rubinfeld, D. L. (1991). *Econometric models and economic forecasts* (3rd ed.). New York: McGraw-Hill.

Pindyck, R. S., & Rubinfeld, D. L. (1998). *Econometric models and economic forecasts* (4th ed.). New York: McGraw-Hill.

Pink, S. (2001). *Doing visual ethnography: Images,*

media and representation in research. London: Sage.

Plackett, R. L. (1972). The discovery of the method of least squares. *Biometrika*, *59*, 239-251.

Plake, B. S., & Impara, J. C. (Eds.). (2001). *The fourteenth mental measurements yearbook.* Lincoln, NE: Buros Institute of Mental Measurements.

Platt, J. (2002). The history of the interview. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research: Context and method* (pp. 33-54). Thousand Oaks, CA: Sage.

Plessy v. Ferguson, 163 U.S. 537 (1896).

Plewis, I. (1985). *Analysing change: Measurement and exploration using longitudinal data.* Chichester, UK: Wiley.

Plummer, K. (2001). *Documents of life 2: An invitation to a critical humanism.* London: Sage.

Poland, B. (2001). Transcription quality. In J. Gubrium & J. Holstein (Eds.), *Handbook of interview research.* Thousand Oaks, CA: Sage.

Poland, B. D. (1995). Transcript quality as an aspect of rigor in qualitative research. *Qualitative Inquiry*, *1*(3), 290-310.

Pollock, K. H., Nichols, J. D., Brownie, C., & Hines, J. E. (1990). Statistical inference for capture-recapture experiments. *Wildlife Monographs*, *107*, 1-97.

Pope, C., & Mayes, N. (1995). Qualitative research: Reaching the parts other methods cannot reach: An introduction to qualitative methods in health and health services research. *British Medical Journal*, *311*, 42-45.

Popkin, S. L. (1991). *The reasoning voter.* Chicago: University of Chicago Press.

Popper, K. (1959). *The logic of scientific discovery.* London: Hutchinson.

Popper, K. (1965). *Conjectures and refutations: The growth of scientific knowledge* (2nd ed.). New York: Basic Books.

Popper, K. (1966). *The open society and its enemies* (5th ed., Vol. 1). London: Routledge. (Original work published 1945)

Popper, K. (1972). *Objective knowledge: An evolutionary approach.* Oxford, UK: Clarendon.

Popper, K. R. (1957). *The poverty of historicism.* Boston: Beacon.

Popper, K. R. (1959). *The logic of scientific discovery.* London: Hutchinson.

Popper, K. R. (1961). *The poverty of historicism.* London: Routledge & Kegan Paul.

Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge.* New York: Basic Books.

Popper, K. R. (1965). *The logic of scientific discovery.* New York: Harper & Row.

Popper, K. R. (1972). *Conjectures and refutation.* London: Routledge & Kegan Paul.

Popper, K. E. (1986). *The poverty of historicism.* London: ARK.

Porter, S. (1993). Critical realist ethnography: The case of racism and professionalism in a medical setting. *Sociology*, *27*, 591-609.

Potter, J. (1996). *Representing reality: Discourse, rhetoric and social construction.* London: Sage.

Potter, J., & Wetherell, M. (1987). *Discourse and social psychology: Beyond attitudes and behaviour.* London: Sage.

Poundstone, W. (1988). *Labyrinths of reason: Paradoxes, puzzles and the frailty of knowledge.* New York: Anchor/ Doubleday.

Poundstone, W. (1992). *Prisoner's dilemma.* New York: Doubleday.

Power, M. (1990). Modernism, postmodernism and organisation. In J. Hassard & D. Pym (Eds.), *The theory and philosophy of organisations.* London: Routledge Kegan Paul.

Powers, D. A., & Xie, Y. (2000). *Statistical methods for categorical data analysis.* San Diego: Academic Press.

Prais, S. J., & Winsten, C. B. (1954). *Trend estimators and serial correlation.* Chicago: Cowles Commission.

Prendergast, C. (1995). *Cultural materialism: On Raymond Williams.* Minneapolis: University of Minnesota Press.

Prentice, R. L., & Zhao, L. P. (1991). Estimating equations for parameters in mean and covariances of multivariate discrete and continuous responses. *Biometrics*, *47*, 825-839.

Preston, S. H., & Coale, A. J. (1982). Age structure, growth, attrition and accession: A new synthesis. *Population Index*, *48*(2), 217-259.

Preston, S. H., Heuveline, P., & Guillot, M.

(2001). *Demography: Measuring and modeling population processes.* Oxford, UK: Basil Blackwell.

Price, L. J. (1989). In the shadow of biomedicine: Selfmedication in two Ecuadorian pharmacies. *Social Science and Medicine*, *28*, 905-915.

Price, L. J. (2002). Carrying out a structured observation. In M. V. Angrosino (Ed.), *Doing cultural anthropology* (pp. 107-114). Prospect Heights, IL: Waveland.

Price, V. (1992). *Public opinion.* Newbury Park, CA: Sage.

Proctor, C. H. (1970). A probabilistic formulation and statistical analysis of Guttman scaling. *Psychometrika*, *35*, 73-78.

Przeworski, A., & Teune, H. (1970). *The logic of comparative social inquiry.* New York: John Wiley.

Quenouille, M. (1949). Approximate tests of correlation in time series. *Journal of the Royal Statistical Society*, *Series B*, *11*, 18-84.

Rabe-Hesketh, S., Pickles, A., & Skrondal, A. (2001). *GLLAMM manual technical report 2001/01.* London: King's College, University of London, Department of Biostatistics and Computing, Institute of Psychiatry. Retrieved from www.iop.kcl.ac.uk/IoP/Departments/Bio Comp/programs/manual.pdf.

Rachlin, H. (2000). *The science of self-control.* Cambridge, MA: Harvard University Press.

Radest, H. B. (1996). *Humanism with a human face: Intimacy and the enlightenment.* London: Praeger.

Radley, A., & Taylor, D. (2003). Remembering one's stay in hospital: A study in recovery, photography and forgetting. *Health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine*, *7*(2), 129-159.

Rae, D. W., & Yates, D. (1981). *Equalities.* Cambridge, MA: Harvard University Press.

Raftery, A. E. (1996). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology* (Vol. 25, pp. 111-163). Oxford, UK: Basil Blackwell.

Raftery, A. E. (1996). Hypothesis testing and model selection. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov chain Monte Carlo in practice* (pp. 163-187). London: Chapman and Hall.

Raghunathan, T. E., & Grizzle, J. E. (1995). A split questionnaire survey design. *Journal of the American Statistical Association*, *90*, 54-63.

Ragin, C. C. (1987). *The comparative method: Moving beyond qualitative and quantitative strategies.* Berkeley: University of California Press.

Ragin, C. C. (1994). *Constructing social research.* Thousand Oaks, CA: Pine Forge Press.

Ragin, C. C. (2000). *Fuzzy-set social science.* Chicago: University of Chicago Press.

Ragin, C. C., & Becker, H. S. (1992). *What is a case? Exploring the foundations of social inquiry.* New York: Cambridge University Press.

Ramazanoglu, C., with Holland, J. (2002). *Feminist methodology.* London: Sage.

Ramsay, J.O. (1977). Maximum likelihood estimation in MDS. *Psychometrika*, *42*, 241-266.

Ramsey, F. P. (1960). Truth and probability. In R. B. Braithwaite (Ed.), *The foundations of mathematics and other logical essays.* New York: Harcourt Brace.

Random House. (1997). *Random House Webster's college dictionary.* New York: Author.

Rao, P., & Griliches, Z. (1969). Small-sample properties of several two-stage regression methods in the context of auto-correlated errors. *Journal of the American Statistical Association*, *64*, 253-272.

Rapoport, A., & Chammah, A. M. (1965). *Prisoner's dilemma.* Ann Arbor: University of Michigan Press.

Rässler, S. (2002). Statistical matching: A frequentist theory, practical applications, and alternative Bayesian approaches. *Lecture Notes in Statistics*, *168*. New York: Springer.

Raudenbush, S. W., & Bryk, A. S. (1988). Methodological advances in analyzing the effects of schools and classrooms on student learning. *Review of Research in Education*, *15*, 423-476.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (1999) *Applied regression analysis: A research tool.* New York: Springer.

Reason, R., & Brabury, H. (Eds.). (2001).

Handbook of action research: Participative inquiry and practice. London: Sage.

Redington, M., & Chater, N. (1997). Probabilistic and distributional approaches to language acquisition. *Trends in the Cognitive Sciences, 1* (7), 273-281.

Reed-Danahay, D. (2001). Autobiography, intimacy and ethnography. In P. Atkinson, A. Coffey, S. Delamont, J. Lofland, & L. Lofland (Eds.), *Handbook of ethnography* (pp. 405-425). London: Sage.

Reed-Danahay, D. E. (Ed.). (1997). *Auto/ethnography: Rewriting the self and the social.* Oxford, UK: Berg.

Rees, P., Martin, D., & Williamson, P. (Eds.). (2002). *The census data system.* Chichester, UK: Wiley.

Reichenbach, H. (1951). *The rise of scientific philosophy.* Berkeley: University of California Press.

Reis, H. T., & Gable, S. L. (2000). Event-sampling and other methods for studying everyday experience. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 190-222). New York: Cambridge University Press.

Rengger, J. (1996). *Retreat from the modern: Humanism, postmodernism and the flight from modernist culture.* Exeter, UK: Bowerdean.

Rescher, N. (1997). *Objectivity: The obligations of impersonal reason.* Notre Dame, IN: University of Notre Dame Press.

Retherford, R. D., & Choe, M. K. (1993). *Statistical models for causal analysis.* New York: John Wiley.

Rex, J. (1974). *Sociology and the demystification of the modern world.* London: Routledge & Kegan Paul.

Reyment, R. A., & Jöreskog, K. G. (1993). *Applied factor analysis in the natural sciences.* Cambridge, UK: Cambridge University Press.

Reynolds, H. T. (1984). *Analysis of nominal data* (2nd ed.). Beverly Hills, CA: Sage.

Rhee, J. W. (1996). How polls drive campaign coverage. *Political Communication, 13*, 213-229.

Rice, J. A. (1995). *Mathematical statistics and data analysis* (2nd ed.). Belmont, CA: Duxbury.

Richards, T. J., & Richards, L. (1994). Using computers in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 445-462). Thousand Oaks, CA: Sage.

Richardson, L. (1985). *The new other woman: Contemporary single women in affairs with married men.* London: Collier-Macmillan.

Richardson, L. (1997). *Fields of play: Constructing an academic life.* New Brunswick, NJ: Rutgers University Press.

Richardson, L. (2000). Writing: A method of inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 923-948). Thousand Oaks, CA: Sage.

Richardson, L., & Lockridge, E. (in press). *Travels with Ernest: Crossing the Literary-Ethnographic Divide.* Walnut Creek, CA: AltaMira.

Ricoeur, P. (1981). *Hermeneutics and the human sciences: Essays on language, action and interpretation.* New York: Cambridge University Press.

Ricoeur, P. (1992). *Oneself as another.* Chicago: The University of Chicago Press.

Ridgeway, C. L. (2001). Inequality, status, and the construction of status beliefs. In J. H. Turner (Ed.), *Handbook of sociological theory* (pp. 323-340). New York: Kluwer Academic/Plenum.

Rieger, J. (1996). Photographing social change. *Visual Sociology, 11*(1), 5-49.

Riessman, C. K. (2002). Doing justice: Positioning the interpreter in narrative work. In W. Patterson (Ed.), *Strategic narrative: New perspectives on the power of personal and cultural storytelling* (pp. 195-216). Lanham, MA: Lexington Books.

Riessman, C. K. (2002). Positioning gender identity in narratives of infertility: South Indian women's lives in context. In M. C. Inhorn & F. van Balen (Eds.), *Infertility around the globe: New thinking on childlessness, gender, and reproductive technologies.* Berkeley: University of California Press.

Riessman, C. K. (2003). Performing identities in illness narrative: Masculinity and multiple sclerosis. *Qualitative Research*, 3(1), 5-33.

Rigdon, E. E. (1995). A necessary and sufficient identification rule for structural models estimated in practice. *Multivariate Behavioral Research*, *30*, 359-383.

Riggs, P. J. (1992). *Whys and ways of science.* Melbourne, Australia: Melbourne University Press.

Riker, W. H. (1980). Implications from the disequilibrium of majority rule for the study of institutions. *American Political Science Review*, *74*, 432-446.

Ripley, B. D. (1996). *Pattern recognition and neural networks.* Cambridge, UK: Cambridge University Press.

Ritchie, J., & Lewis, J. (2003). *Qualitative research practice.* London: Sage.

Ritzer, G. (1980). *Sociology: A multiple paradigm science.* Boston: Allyn & Bacon.

Rivers, D., & Vuong, Q. H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics*, *39*, 347-366.

Robert, C. P. (2001). *The Bayesian choice: A decision theoretic motivation* (2nd ed.). New York: Springer-Verlag.

Robert, C. P., & Casella, G. (1999). *Monte Carlo statistical methods.* New York: Springer-Verlag.

Roberts, B. (2002). *Biographical research.* Buckingham, UK: Open University Press.

Roberts, N., Andersen, D. F., Deal, R. M., Grant, M. S., & Shaffer, W. A. (1983). *Introduction to computer simulation: A system dynamics modeling approach.* Reading, MA: Addison-Wesley.

Robey, B., Rutstein, S. O., & Morris, L. (1992). The reproductive revolution: New survey findings (Technical Report M-11). *Population Reports.* Baltimore, MD: Johns Hopkins University Center for Communication Programs.

Robins, J. M. (1998). Correction for non-compliance in equivalence trials. *Statistics in Medicine*, *17*, 269-302.

Robinson, J. (1994). White women researching/representing "others": From anti-apartheid to postcolonialism. In A. Blunt & G. Rose (Eds.), *Writing, women and space* (pp. 97-226). New York: Guilford.

Robinson, J. P., & Godbey, G. (1997). *Time for life.* University Park: Pennsylvania State University Press.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*, 351-357.

Robinson, W. S. (1951). The logical structure of analytic induction. *American Sociological Review*, *16*, 812-818.

Robson, C. (1993). *Real world research.* Oxford, UK: Blackwell.

Rochford, E. B., Jr. (1992). On the politics of member validation: Taking findings back to Hare Krishna. In G. Miller & J. A. Holstein (Eds.), *Perspectives on social problems* (Vol. 3, pp. 99-116). Greenwich, CT: JAI.

Rodgers, J. L. (1999). The bootstrap, the jackknife, and the randomization test: A sampling taxonomy. *Multivariate Behavioral Research*, *34*, 441-456.

Rodríguez, G. (1994). Statistical issues in the analysis of reproductive histories using hazard models. In K. L. Campbell & J. W. Wood (Eds.), *Human reproductive ecology: Interaction of environment, fertility and behavior* (pp. 266-279). New York: New York Academy of Sciences.

Roethlisberger, F. J., & Dickson, W. J. (1939). *Management and the worker.* Cambridge, MA: Harvard University Press.

Rogers, A. (1995). *Multiregional demography.* Chichester, UK: Wiley.

Rohrer, L. G. (1965). The great response-style myth. *Psychological Bulletin*, *63*, 129-156.

Romney, A. K., Weller, S. C., & Batchelder, W. (1986). Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist*, *88*, 313-338.

Ronan, W. W., & Latham, G. P. (1974). The reliability and validity of the critical incident technique: A closer look. *Studies in Personnel Psychology*, *6*(1), 33-64.

Roncek, D. W. (1992). Learning more from Tobit coefficients. *American Sociological Review*, *57*, 503-507.

Rorty, R. (1991). *Objectivity, relativism, and truth: Philosophical papers volume 1.* Cambridge, UK: Cambridge University Press.

Rosaldo, R. (1989). *Culture & truth.* Boston:

Beacon.

Rose, G. (2001). *Visual methodologies*. London: Sage.

Rosenau, P. M. (1992). *Post-modernism and the social sciences: Insights, inroads and intrusions*. Princeton, NJ: Princeton University Press.

Rosenbaum, P. R. (1995). *Observational studies*. New York: Springer-Verlag.

Rosenbaum, P. R. (1998). Multivariate matching methods. *Encyclopedia of Statistical Sciences*, *2*, 435-438.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41-55.

Rosenberg, A. (1988). *Philosophy of social science*. Oxford, UK: Clarendon.

Rosenberg, M. (1968). *The logic of survey analysis*. New York: Basic Books.

Rosenberg, M. J. (1969). The conditions and consequences of evaluation apprehension. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 279-349). New York: Academic Press.

Rosenberg, S., & Kim, M. P. (1975). The method of sorting as a data-gathering method in multivariate research. *Multivariate Behavioral Research*, *10*, 489-502.

Rosenberger, J. L., & Gasko, M. (1983). Comparing location estimators: Trimmed means, medians, and trimean. In D. C. Hoaglin, F. Mosteller, & J.W. Tukey (Eds.), *Understanding robust and exploratory data analysis*. New York: Wiley.

Rosenhan, D. (1973). On being sane in insane places. *Science*, *179*, 250-258.

Rosenthal, G. (Ed.). (1998). *The Holocaust in three generations: Families of victims and persecutors of the Nazi regime*. London: Cassell.

Rosenthal, R. (1966). *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts.

Rosenthal, R. (1976). *Experimenter effects in behavioral research* (rev. ed.). New York: Irvington.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, *86*, 638-641.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart & Winston.

Rosenthal, R., & Jacobson, L. (1992). *Pygmalion in the classroom: Expanded edition*. New York: Irvington.

Rosenthal, R., & Rosnow, R. L. (1975). *The volunteer subject*. New York: Wiley.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.

Rosenthal, R., & Rosnow, R. L. (Eds.). (1969). *Artifact in behavioral research*. New York: Academic Press.

Rosenthal, R., & Rubin, D. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences*, *3*, 377-415.

Rosenzweig, M. R., & Wolpin, K. I. (2000). Natural "natural experiments" in economics. *Journal of Economic Literature*, *38*, 827-874.

Rosenzweig, S. (1933). The experimental situation as a psychological problem. *Psychological Review*, *40*, 337-354.

Rosnow, R. L., & Rosenthal, R. (1970). Volunteer effects in behavioral research. In K. H. Craik, B. Kleinmuntz, R. L. Rosnow, R. Rosenthal, J. A. Cheyne, & R.H.Walters (Eds.), *New directions in psychology* (pp. 213-277). New York: Holt, Rinehart and Winston.

Rosnow, R. L., & Rosenthal, R. (1997). *People studying people: Artifacts and ethics in behavioral research*. New York: W. H. Freeman.

Ross, S. (1999). *A first course in probability*. Upper Saddle River, NJ: Prentice Hall.

Rossi, P. H. (1951). *The application of latent structure analysis to the study of social stratification*. Unpublished doctoral dissertation, Columbia University.

Rossi, P. H. (1979). Vignette analysis: Uncovering the normative structure of complex judgments. In R. K. Merton, J. S. Coleman, & P. H. Rossi (Eds.), *Qualitative and quantitative social research: Papers in honor of Paul F. Lazarsfeld* (pp. 176-186). New York: Free Press.

Rossi, P. H., & Anderson, A. B. (1982). The factorial survey approach: An introduction. In P. H.

Rossi & S. L. Nock (Eds.), *Measuring social judgments: The factorial survey approach* (pp. 15-67). Beverly Hills, CA: Sage.

Rossi, P. H., & Berk, R. A. (1985). Varieties of normative consensus. *American Sociological Review*, *50*, 333-347.

Rossi, P. H., Freeman, H. E., & Lipsey, M. (1998). *Evaluation: A systematic approach* (6th ed.). Thousand Oaks, CA: Sage.

Rossi, P. H., Sampson, W. A., Bose, C. E., Jasso, G., & Passel, J. (1974). Measuring household social standing. *Social Science Research*, *3*, 169-190.

Rothenberg, J., & Rothenberg, D. (Eds.). (1983). *Symposium of the whole: A range of discourse toward an ethnopoetics.* Berkeley: University of California Press.

Rothman, K. J. (1977). Epidemiologic methods in clinical trials. *Cancer*, *39*, 1771-1775.

Rothman, K. J. (1986). *Modern epidemiology.* Boston: Little, Brown.

Rothman, K. J., & Greenland, S. (1998). *Modern epidemiology* (2nd ed.). Philadelphia: Lippincott.

Rousseau, J.-J. (1953) *The confessions.* Harmondsworth, UK: Penguin. (Original work published 1781)

Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regression & outlier detection.* New York: Wiley.

Routledge, P. (1996). The third space as critical engagement. *Antipode*, *28*(4), 399-419.

Rowell, R. K. (1996). Partitioning predicted variance into constituent parts: How to conduct commonality analysis. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 4, pp. 33-44). Greenwich, CT: JAI.

Royston, J. P. (1993). A toolkit for testing for non-normality in complete and censored samples. *The Statistician*, *42*, 37-43.

Rubenstein, A. (1998). *Modeling bounded rationality.* Cambridge: MIT Press.

Rubin, D. B. (1986). Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business and Economic Statistics*, *4*, 87-95.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* New York: Wiley.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, *2*, 169-188.

Rubin, D. B., & Schenker, N. (1986). Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association*, *81*, 366-374.

Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo method.* New York: John Wiley.

Rudas, T. (1997). *Odds ratios in the analysis of contingency tables.* Thousand Oaks, CA: Sage.

Rummel, R. J. (1970). *Applied factor analysis.* Evanston, IL: Northwestern University Press.

Ruskey, F. (2001). A survey of Venn diagrams. *The Electronic Journal of Combinatorics*, Dynamic Survey #5 [Online]. Retrieved from http://www.combinatorics.org/Surveys/ds5/VennEJC.html.

Russell, B. (1912). *Mysticism and logic.* London: Allen & Unwin.

Russett, B. (1969). Inequality and instability: The relation of land tenure to politics. In D. Rowney & J. Graham (Eds.), *Quantitative history: Selected readings in the quantitative analysis of historical data* (pp. 356-367). Homewood, IL: Dorsey.

Ryle, G. (1971). Thinking and reflecting. In G. Ryle, *Collected papers, Volume 2.* London: Hutchinson.

Ryssevik, J., & Musgrave, S. (2001). The social science dream machine: Resource discovery, analysis, and delivery on the Web. *Social Science Computing Review*, *19*(2), 163-174.

Sackett, G. P. (1979). The lag sequential analysis of contingency and cyclicity in behavioral interaction research. In J. D. Osofsky (Ed.), *Handbook of infant development* (1st ed., pp. 623-649). New York: Wiley.

Sacks, H. (1992). *Lectures on conversation* (2 vols.). Oxford, UK: Blackwell.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696-735.

Sade, A. (1951). An omission in Norton's list of 7×7 squares. *Annals of Mathematical Statistics*, *22*,

306-307.

Sale, J. E. M., Lohfeld, L. H., & Brazil, K. (2002). Revisiting the quantitative-qualitative debate: Implications for mixedmethods research. *Quality and Quantity*, *36*, 43-53.

Salem, A., & Mount, T. (1974). A convenient descriptive model of income distribution: The gamma density. *Econometrica*, *42*, 1115-1127.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world.* Princeton, NJ: Princeton University Press.

Salsburg, D. (2001). *The lady tasting tea: How statistics revolutionized science in the twentieth century.* New York: Holt.

Sandelowski, M., Docherty, S., & Emden, C. (1997). Qualitative metasynthesis: Issues and techniques. *Research in Nursing & Health*, *20*, 365-371.

Sankoff, D., & Kruskal, J. B. (1983). *Time warps, string edits, and macromolecules.* Reading, MA: Addison-Wesley.

Sapir, E. (1951). Culture, genuine and spurious. In D. G. Mandelbaum (Ed.), *Selected writings of Edward Sapir in language, culture and personality* (pp. 308-331). Berkeley: University of California Press.

Saris, W. E. (1991). *Computer-assisted interviewing.* Newbury Park, CA: Sage.

Saris, W. E., Satorra, A., & Sörbom, D. (1987). The detection and correction of specification errors in structural equation models. In C. C. Clogg (Ed.), *Sociological methodology 1987* (pp. 105-129). San Francisco: Jossey-Bass.

Särndal, C. E., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling.* New York: Springer-Verlag.

Saunders, M. N. K., & Cooper, S. A. (1993). *Understanding business statistics: An active learning approach.* London: DP Publications.

Saunders, P. (1979). *Urban politics: A sociological interpretation.* London: Hutchinson.

Saunders, P. T. (1980). *An introduction to catastrophe theory.* New York: Cambridge University Press.

Saussure, F. (1966). *Course in general linguistics.* New York: McGraw-Hill. (Originally published in 1915)

Saussure, F. de (1974). *Course in general linguistics.* London: Fontana/Collins.

Sayer, A. (1992). *Method in social science: A realist approach* (2nd ed.). London: Routledge.

Sayer, A. (1997). Essentialism, social constructionism and beyond. *Sociological Review*, *45*, 453-487.

Sayer, A. (2000). *Realism and social science.* London: Sage.

Sayer, A. (2000). System, lifeworld and gender: Associational versus counterfactual thinking. *Sociology*, *34*, 707-725.

Sayer, A. (2001). Reply to Holmwood. *Sociology*, *35*, 967-984.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* London: Chapman & Hall.

Scheaffer, R. L., Mendenhall, W., & Ott, L. (1990) *Elementary survey sampling.* Boston: PWS-Kent.

Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika*, *40*, 87-104.

Scheffé, H. (1959). *The analysis of variance.* New York: Wiley.

Schegloff, E. A. (1968). Sequencing in conversational openings. *American Anthropologist*, *70*, 1075-1095.

Schegloff, E. A. (1989). Harvey Sacks—lectures 1964-1965: An introduction/memoir. *Human Studies*, *12*, 185-209.

Schegloff, E. A. (1992). Repair after next turn: The last structurally provided defence of intersubjectivity in conversation. *American Journal of Sociology*, *97* (5), 1295-1345.

Schegloff, E. A. (1996). Confirming allusions: Toward an empirical account of action. *American Journal of Sociology*, *2*, 161-216.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, *1*, 143-186.

Schelling, T. C. (1978). *Micromotives and macrobehavior.* New York: Norton.

Schensul, J. J., & LeCompte, M. D. (Eds.). (1999). *Ethnographer's toolkit.* Walnut Creek, CA: AltaMira.

Schensul, S. L., Schensul, J. J., & LeCompte, M.

D. (1999). *Essential ethnographic methods: Observations, interviews, and questionnaires.* Walnut Creek, CA: AltaMira.

Schervish, M. J. (1996). *P*-values: What they are and what are they not. *American Statistician*, *50*, 203-206.

Schlesselman, J. J. (1982). *Case-control studies: Design, conduct, analysis.* New York: Oxford University Press.

Schmidt, P. (1976). *Econometrics.* New York: Marcel Dekker.

Schmidt, P., & Witte, A. D. (1988). *Predicting recidivism using survival models.* New York: Springer-Verlag.

Schneider, H. (1986). *Truncated and censored samples from normal populations.* New York: Marcel Dekker.

Schober, M., & Conrad, F. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly*, *61*, 576-602.

Schoenberg, R. (1977). Dynamic models and cross-sectional data: The consequences of dynamic misspecification. *Social Science Research*, *6*, 133-144.

Schölkopf, B., & Smola, A. J. (2003). *Learning with kernel.* Cambridge: MIT Press.

Schrodt, P. A. (2002). Mathematical modeling. In J. B. Manheim, R. C. Rich, & L. Willnat (Eds.), *Empirical political analysis: Research methods in political science* (5th ed.). New York: Longman.

Schroeder, L. D., Sjoquist, D. L., & Stephan, P. E. . (1986). *Understanding regression analysis: An introductory guide.* Beverly Hills, CA: Sage.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys.* New York: Academic Press.

Schuman, H., & Presser, S. (1996). *Questions & answers in attitude surveys.* Thousand Oaks, CA: Sage.

Schütz, A. (1963). Concept and theory formation in the social sciences. In M. A. Natanson (Ed.), *Philosophy of the social sciences* (pp. 231-249). New York: Random House.

Schutz, A. (1964). Don Quixote and the problem of reality. In *Collected papers*, *Vol. 2.* The Hague: Martinus Nijhoff.

Schütz, A. (1967). *Der sinnhafte Aufbau der sozialen Welt* [Phenomenology of the social world]. Evanston, IL: Northwestern University Press. (Originally published in 1932)

Schutz, A. (1973). *Collected papers* (4 vols.). Dordrecht, The Netherlands: Kluwer.

Schütz, A. (1976). Equality and the meaning structure of the social world. In *Collected Papers II* (pp. 226-273). The Hague: Martinus Nijhoff. (Originally published in 1955)

Schütz, A. (1976). *The phenomenology of the social world* (G. Walsh & F. Lehnert, Trans.). London: Heinemann.

Schwandt, T. (1996). Farewell to criteriology. *Qualitative Inquiry*, *2*, 58-72.

Schwandt, T. A. (1998). The interpretive review of educational matters: Is there any other kind? *Review of Educational Research*, *68*(4), 409-412.

Schwarz, N., & Sudman, S. (Eds.). (1996). *Answering questions: Methodology for determining cognitive and communicative processes in survey research.* San Francisco: Jossey-Bass.

Schwarz, N., Hippler, H.-J., Deutsch, B., & Strack, F. (1985). Response scales: Effects of category range on reported behavior and comparative judgments. *Public Opinion Quarterly*, *49*(3), 388-395.

Scott, J. (1990). *A matter of record: Documentary sources in social research.* Cambridge, UK: Polity.

Scott, J. (1992). *Social network analysis.* London: Sage.

Scott, J. W. (1988). *Gender and the politics of history.* New York: Columbia University Press.

Scott, M. B., & Lyman, S. (1968). Accounts. *American Sociological Review*, *33*, 46-62.

Scott, W. A. (1955). Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly*, *19*, 321-325.

Scott, W. A. (1968). Attitude measurement. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology.* Reading, MA: Addison-Wesley.

Scriven, M. (1959). Explanation and prediction in evolutionary theory. *Science*, *130*, 477-482.

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.

Scriven, M. (2001). Evaluation: Future tense. *The*

American Journal of Evaluation, *22*, 301-307.

Seal, H. L. (1967). The historical development of the Gauss linear model. *Biometrika*, *54*, 1-23.

Seale, C. F. (1999). *The quality of qualitative research*. London: Sage.

Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components*. New York: Wiley.

Sechrest, L. (Ed.). (1979). *Unobtrusive measures today*. San Francisco: Jossey-Bass.

Segrè, E. (1970). *Enrico Fermi: Physicist*. Chicago: University of Chicago Press.

Sen, A. (1973). *On economic inequality*. New York: Norton.

Sewell, W. H. (1942). The development of a sociometric scale. *Soociometry*, *5*(3), 279-297.

Shadish, W. R. (2000). The empirical program of quasiexperimentation. In L. Bickman (Ed.), *Validity and social experimentation: Donald Campbell's legacy* (pp. 13-35). Thousand Oaks, CA: Sage.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin.

Shadish, W. R., Jr., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage.

Shadish, W. R., Jr., Newman, D. L., Scheirer, M. A., & Wye, C. (1995). *Guiding principles for evaluators* (New Directions for Program Evaluation, No. 66). San Francisco: Jossey-Bass.

Shapiro, G., & Markoff, J. (1998). *Revolutionary demands: A content analysis of the Cahier de Doléances of 1789*. Stanford, CA: Stanford University Press.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*, 591-611.

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.

Shaw, C. (1966). *The jack roller*. Chicago: University of Chicago Press.

Shaw, I., Bloor, M., Cormack, R., & Williamson, H. (1996). Estimating the prevalence of hard-to-reach populations: The illustration of mark-recapture methods in the study of homelessness. *Social Policy and Administration*, *30*(1), 69-85.

Sheskin, D. J. (1997). *Handbook of parametric and nonparametric statistical procedures*. Boca Raton, FL: CRC Press.

Shneiderman, B. (1992). Tree visualization with treemaps: A 2-D space-filling approach. *ACM Transactions on Graphics*, *11*(1), 92-99.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlation: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420-428.

Shryock, H. S., Siegel, J. S., & Associates. (1973). *The methods and materials of demography*. New York: Academic Press. (Condensed edition by Edward G. Stockwell.)

Si, M., Neufeld, R. R., & Dunbar, J. (1999). Removal of bedrails on a short-term nursing home rehabilitation unit. *The Gerontologist*, *39*, 611-614.

Sieber, J. E. (1992). *Planning ethically responsible research*. Newbury Park, CA: Sage.

Sieber, J. E. (1996). Typically unexamined communication processes in research. In B. H. Stanley, J. E. Sieber, & G. B. Melton (Eds.), *Research ethics: A psychological approach*. Lincoln: University of Nebraska Press.

Sieber, J. E. (Ed.). (1991). *Sharing social science data: Advantages and challenges*. Newbury Park, CA: Sage.

Siegel, J. S. (2002). *Applied demography*. San Diego, CA: Academic Press.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.

Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory*. Thousand Oaks, CA: Sage.

Silverman, D. (1993). *Interpreting qualitative data: Methods for analysing talk, text, and interaction*. London: Sage.

Simon, H. A. (1954). Spurious correlation: A causal interpretation. *Journal of the American Statistical Association*, *49*, 467-479.

Simon, H. A. (1957). *Models of man*. New York: Wiley.

Simon, H. A. (1976). Discussion: Cognition and social behavior. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior.* Hillsdale, NJ: Erlbaum.

Simons, H. (1996). The paradox of case study. *Cambridge Journal of Education*, *26*(2), 225-240.

Sims, C. A. (1972). Money income and causality. *American Economic Review*, *62*, 540-552.

Singer, E. (1978). Informed consent: Consequences for response rate and response quality in social surveys. *American Sociological Review*, *43*, 144-162.

Singer, E., & Frankel, M. R. (1982). Informed consent procedures in telephone interviews. *American Sociological Review*, *47*, 416-427.

Singer, E., vonThurn, D., & Miller, E. (1995). Confidentiality assurances and response. *Public Opinion Quarterly*, *59*, 66-77.

Singleton, A. (1999). Measuring international migration: A case study of European cross-national comparisons. In D. Dorling & S. Simpson (Eds.), *Statistics in society: The arithmetic of politics* (pp. 148-158). London: Arnold.

Skinner, C. J., Holt, D., & Smith, T. M. F. (Eds.). (1989). *Analysis of complex surveys.* Chichester, UK: Wiley.

Sklodowska, E. (1992). *Testimonio hispanoamericano: Historia, teoría, poética.* New York: Peter Lang.

Skoudis, E. (2002). *Counter Hack: A step-by-step guide to computer attacks and effective defenses.* Upper Saddle, NJ: Prentice Hall.

Skvoretz, J., & Faust, K. (1999). Logit models for affiliation networks. *Sociological Methodology*, *29*, 253-280.

Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher*, *15*(9), 5-11.

Smart, B. (2000). Postmodern social theory. In B. Turner (Ed.), *The Blackwell companion to social theory* (pp. 447-480). Oxford, UK: Blackwell.

Smelser, N. (1997). *Problems of sociology.* Berkeley: University of California Press.

Smith, B. H. (1997). *Belief and resistance: Dynamics of contemporary intellectual controversy.* Cambridge, MA: Harvard University Press.

Smith, D. (1997). Comment on Heckman's "Truth and method: Feminist standpoint theory revisited." *Signs*, *22*(21): 392-397.

Smith, D. P. (1992). *Formal demography.* New York: Plenum.

Smith, G. (1997, Winter). Do statistics test scores regress toward the mean? *Chance*, pp. 42-45.

Smith, J. (1993). *After the demise of empiricism: The problem of judging social and educational inquiry.* Norwood, NJ: Ablex.

Smith, J. K. (1989). *The nature of social and educational inquiry: Empiricism versus interpretation.* Norwood, NJ: Ablex.

Smith, J. K., & Heshusius, L. (1986). Closing down the conversation: The end of the quantitative-qualitative debate among educational enquirers. *Educational Researcher*, *15*, 4-12.

Smith, J., & Deemer, D. (2000). The problem of criteria in the age of relativism. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 877-922). Thousand Oaks, CA: Sage.

Smith, L. M. (1978). An evolving logic of participant observation, educational ethnography and other case studies. *Review of Research in Education*, *6*, 316-377.

Smith, M. (1998). *Social science in question.* London: Sage.

Smith, N. W. (2001). *Current systems in psychology.* Belmont, CA: Wadsworth.

Smith, S. K., Tayman, J., & Swanson, D. A. (2001). *State and local population projections: Methodology and analysis.* New York: Kluwer Academic/Plenum Publishers.

Smith, S., & Watson, J. (2001). *Reading autobiography: A guide for interpreting life narratives.* Minneapolis: University of Minnesota Press.

Smith, T. W. (1992). Thoughts on the nature of context effects. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 163-184). NewYork: Springer-Verlag.

Smithson, M. (1987). *Fuzzy set analysis for behavioral and social sciences.* New York: Springer-Verlag.

Smithson, M. (2003). *Confidence intervals* (Sage University Papers on Quantitative Applications in the Social Sciences, 07-140). Thousand Oaks, CA: Sage.

Sneath, P. H. A., & Sokal, R. R. (1973). *Numerical taxonomy: The principles and practice of numerical classification.* San Francisco: Freeman.

Snedecor, G.W. (1937). *Statistical methods.* Ames: Iowa State University Press.

Snedecor, G. W. (1956). *Statistical methods* (5th ed.). Ames: Iowa State College Press.

Snider, J. G., & Osgood, C. E. (Eds.). (1969). *Semantic differential technique: A sourcebook.* Chicago: Aldine.

Snijders, T. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology*, *31*, 361-395.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London: Sage.

Snow, D. (1980). The disengagement process: A neglected problem in participant observation research. *Qualitative Sociology*, *3*, 100-122.

Sobel, M. E. (1995). Causal inference in the social and behavioral sciences. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 1-38). New York: Plenum.

Sobel, M., & Bohrnstedt, G. W. (1985). Use of null models in evaluating the fit of covariance structure models. In N. B. Tuma (Ed.), *Sociological methodology 1985* (pp. 152-178). San Francisco: Jossey-Bass.

Sokal, R. R., & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, *38*, 1409-1438.

Sokal, R. R., & Rohlf, F. J. (1981). *Biometry* (2nd ed.). San Francisco: W.H. Freeman.

Sokal, R., & Sneath, P. (1963). *Principles of taxonomy.* San Francisco: Freeman.

Solomon, R. L. (1949). An extension of control group design. *Psychological Bulletin*, *46*, 137-150.

Solomon, W. D. (1978). Ethics: Rules and principles. In W. T. Reich (Ed.), *Encyclopedia of bioethics.* New York: Free Press.

Sørensen, A. B. (1979). A model and a metric for the analysis of the intragenerational status attainment process. *American Journal of Sociology*, *85*, 361-384.

Sørensen, J. B. (2002). The use and misuse of the coefficient of variation in organizational demography research. *Sociological Methods & Research*, *30*, 475-491.

Soulliere, D., Britt, D. W., & Maines, D. R. (2001). Conceptual modeling as a toolbox for grounded theorists. *Sociological Quarterly*, *42*, 233-251.

Soyland, A. J. (1994). *Psychology as metaphor.* London: Sage.

Spanos, A. (1999). *Probability and statistical inference: Econometric modeling with observational data.* Cambridge, UK: Cambridge University Press.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, *15*, 366-374.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, *15*, 72-101.

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, *12*, 271-295.

Special section: Speaking in the name of the real: Freeman and Mead on Samoa. (1983). *American Anthropologist*, *85*, 908-947.

Spector, P. E. (1987). Method variance as an artifact in selfreported affect and perceptions at work: Myth or significant problem? *Journal of Applied Psychology*, *72*, 438-443.

Spector, P. E. (1992). *Summated rating scale construction: An introduction* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-082). Newbury Park, CA: Sage.

Spector, P. E. (1994). Using self-report questionnaires in OB research: A comment on the use of a controversial method. *Journal of Organizational Behavior*, *15*, 385-392.

Spector, P. E., & Brannick, M. T. (1995). The nature and effects of method variance in organizational research. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational*

psychology: 1995 (pp. 249-274). West Sussex, England: Wiley.

Spector, P. E., Van Katwyk, P. T., Brannick, M. T., & Chen, P.Y. (1997). When two factors don't reflect two constructs: How item characteristics can produce artifactual factors. *Journal of Management*, *23*, 659-677.

Spiegel, M. R., Schiller, J., & Srinivasan, R. A. (2000). *Schaum's outline of probability and statistics* (2nd ed.). New York: McGraw-Hill.

Spradley, J. P. (1979). *The ethnographic interview.* New York: Holt, Rinehart and Winston.

SPSS base 10.0 user's guide. (1999). Chicago: SPSS.

SPSS. (1990). *SPSS reference guide*. Chicago: SPSS.

SPSS. (2001). *Reliability* (Technical section). Retrieved from www.spss.com/tech/stat/algorithms/11.0/reliability.pdf.

SPSS. (2002, August). Crosstabs. In *SPSS 11.0 statistical algorithms* [Online]. Available: http://www.spss.com/tech/stat/algorithms/11.0/crosstabs.pdf.

Stanley, L. (1992). *The auto/biographical I: Theory and practice of feminist auto/biography.* Manchester, UK: Manchester University Press.

Stanley, L., & Wise, S. (1993). *Breaking out again: Feminist ontology and epistemology.* London: Routledge Kegan Paul.

Stanton, A. L., Burker, E. J., & Kershaw, D. (1991). Effects of researcher follow-up of distressed subjects: Tradeoff between validity and ethical responsibility? *Ethics & Behavior*, *1*(2), 105-112.

Staudte, R. G., & Sheather, S. J. (1990). *Robust estimation and testing.* New York: Wiley.

Stem, D. E., Jr., & Lamb, C. W., Jr. (1981). The marble-drop technique: A procedure for gathering sensitive information. *Decision Sciences*, *12*, 702-707.

Stern, J., Stackowiack, R., & Greenwald, R. (2001). *Oracle essentials: Oracle9i, Oracle8i and Oracle 8.* Sebastopol, CA: O'Reilly.

Stevens, J. (2001). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*, 677-680.

Stevens, S. S. (1951). *Handbook of experimental psychology.* New York: Wiley.

Stevens, S. S. (1968). Measurement, statistics, and the schemapiric view. *Science*, *161*, 849-861.

Stewart, D. K., & Love, W. A. (1968). A general canonical correlation index. *Psychological Bulletin*, *70*, 160-163.

Stewart, D. W., & Shamdasani, P. N. (1990). *Focus groups: Theory and practice* (Applied Social Research Methods Series, Vol. 20). Newbury Park, CA: Sage.

Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900.* Cambridge, MA: Belknap Press of Harvard University Press.

Stigler, S. (1986). Laplace's 1774 memoir on inverse probability. *Statistical Science*, *1*, 359-378.

Stigler, S. M. (1999). *Statistics on the table: The history of statistical concepts and methods.* Cambridge, MA: Harvard University Press.

Stoll, D. (1999). *Rigoberta Menchú and the story of all poor Guatemalans.* Boulder, CO: Westview.

Stone, A. A., Shiffman, S., & DeVries, M. (1999). Ecological momentary assessment. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 27-38). New York: Russell Sage.

Stone, A. A., Turkkan, J. S., Bachrach, C. A., Jobe, J. B., Kurtzman, H. S., & Cain, V. S. (Eds.). (2000). *The science of selfreport: Implications for research and practice.* Mahwah, NJ: Lawrence Erlbaum.

Stouffer, S. A. (1950). Some observations on study design. *American Journal of Sociology*, *55*, 355-361.

Strauss, A. (1987). *Qualitative analysis for social scientists.* New York: Cambridge University Press.

Strauss, A. L., Corbin, J., Fagerhaugh, S., Glaser, B., Maines, D., Suczek, B., & Weiner, C. (1984). *Chronic illness and the quality of life.* St. Louis, MO: Mosby.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury Park, CA: Sage.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Grounded theory procedures and techniques* (2nd ed.). Thousand Oaks, CA: Sage.

Stringer, E. T. (1996). *Action Research: A handbook for practitioners*. Thousand Oaks, CA: Sage.

Strohmetz, D. B., & Rosnow, R. L. (1994). A mediational model of research artifacts. In J. Brzezinski (Ed.), *Probability in theory-building: Experimental and nonexperimental approaches to scientific research in psychology* (pp. 177-196). Amsterdam: Editions Rodopi.

Stuart, A. (1984). *The ideas of sampling*. London: Griffin.

Stuart, A., & Ord, J. K. (1987). *Kendall's advanced theory of statistics* (Vol. 1). London: Edward Arnold.

Stuart, A., Ord, J. K., & Arnold, S. (1999). *Kendall's advanced theory of statistics: Vol. 2A. Classical inference & the linear model* (6th ed.). New York: Oxford University Press.

Studenmund, A. H. (1992). *Using econometrics: A practical guide* (2nd ed.). New York: HarperCollins.

Student. (1908). The probable error of a mean. In E. S. Pearson & J. Wishart (Eds.), *Student's collected papers*. London: University College.

Suchman, E. A. (1967). *Evaluative research: Principles and practice in public service and social action programs*. New York: Russell Sage.

Sudman, S. (1976). *Applied sampling*. New York: Academic Press.

Sudman, S., & Bradburn, N. (1982). *Asking questions*. San Francisco: Jossey-Bass.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.

Sudnow, D. (1978). *Ways of the hand*. Cambridge, MA: Harvard University Press.

Sudnow, D. (2001). *Ways of the hand* (Rev. ed.). Cambridge, MA: MIT Press.

Suen, H. K., & Ary, D. (1989). *Analyzing quantitative behavioural observation data*. London: Lawrence Erlbaum.

Suppes, P. (1970). *A probabilistic theory of causality*. Amsterdam: North Holland.

Swokowski, E. W. (1979). *Calculus with analytic geometry* (2nd ed.). Boston: Prindle, Weber & Schmidt.

Symon, G. J., & Clegg, C. W. (1991). Technology-led change: A study of the implementation of CADCAM. *Journal of Occupational Psychology*, *64*, 273-290.

Szalai, A. (Ed.). (1972). *The use of time*. The Hague: Mouton.

Sztompka, P. (1990). *Robert Merton: An intellectual profile*. London: Macmillan.

Tabachnick, B. G., & Fidell, L. S. (2001). *Computer-assisted research design and analysis*. Boston: Allyn and Bacon.

Tacq, J. J. A. (1984). *Causaliteit in Sociologisch Onderzoek. Een Beoordeling van Causale Analysetechnieken in het Licht van Wijsgerige Opvattingen over Causaliteit* [Causality in sociological research: An evaluation of causal techniques of analysis in the light of philosophical theories of causality]. Deventer, The Netherlands: Van Loghum Slaterus.

Tacq, J. J. A. (1997). *Multivariate analysis techniques in social science research: From problem to analysis*. London: Sage.

Takane, Y. (1981). Multidimensional scaling of sorting data. In Y. P. Chaubey & T. D. Dwivedi (Eds.), *Topics in applied statistics*. Montreal: Concordia University Press.

Takane, Y., Young, F. W., & DeLeeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, *42*, 7-67.

Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association*, *82*, 528-550.

Taris, T. W. (2000). *A primer in longitudinal data analysis*. London: Sage.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.

Taylor, C. (1994). The politics of recognition. In C. Taylor & A. Gutmann (Eds.), *Multiculturalism and the politics of recognition* (pp. 25-73). Princeton, NJ: Princeton University Press.

Taylor, M. F., with Brice, J., Buck, N., & Prentice-Lane, E. (Eds.). (2000). *British Household Panel Survey user manual volume B9: Codebook*.

Colchester, UK: University of Essex.

Taylor, S. J., & Bogdan, R. (1998). *Introduction to qualitative research methods: A guide & resource.* New York: Wiley.

Tedlock, D. (1999). Poetry and ethnography: A dialogical approach. *Anthropology and Humanism*, *24*(2), 155-167.

ten Have, P. (1999). *Doing conversation analysis.* Thousand Oaks, CA: Sage.

Tenenhaus, M. (1998). *La régression PLS* [PLS regression]. Paris: Technip.

The 1980 U.S. census [Special issue]. (1992). *Survey Methodology*, *18*.

The 1990 U.S. census [Special issue]. (1993). *Journal of the American Statistical Association*, *88*.

The 1990 U.S. census [Special issue]. (1994). *Statistical Science*, *9*.

The 2000 U.S. census. (2001). *Society*, *39*, 2-53.

Therneau, T. M., & Grambsch, P. M. (2000). *Modeling survival data: Extending the Cox model.* New York: Springer-Verlag.

Theus, M. (1997). Visualization of categorical data. In W. Bandilla & F. Faulbaum (Eds.), *SoftStat '97: Advances in statistical software* (Vol. 6, pp. 47-55). Stuttgart, Germany: Lucius & Lucius.

Thom, R. (1975). *Structural stability and morphogenesis.* Reading, MA: Benjamin.

Thomas, H. (1997). Dancing: Representation and difference. In J. McGuigan (Ed.), *Cultural methodologies.* London: Sage.

Thomas, J. (1992). *Doing critical ethnography.* Newbury Park, CA: Sage.

Thomas, W. I., & Znaniecki, F. (1918-1920). *The Polish peasant in Europe and America* (5 vols.). New York: Dover.

Thomas, W. I., & Znaniecki, F. (1958). *The Polish peasant in Europe and America* (2 vols.). New York: Dover. (Original five-volume work published in 1918-1920.)

Thompson, B. (1985). Alternate methods for analyzing data from experiments. *Journal of Experimental Education*, *54*, 50-55.

Thompson, B. (1997). The importance of structure coefficients in structural equation modeling confirmatory factor analysis. *Educational and Psychological Measurement*, *57*, 5-19.

Thompson, B. (2000). Canonical correlation analysis. In L. Grimm & P. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 285-316). Washington, DC: American Psychological Association.

Thompson, J. (1981). *Critical hermeneutics: A study in the thought of Paul Ricoeur and Jürgen Habermas.* Cambridge, UK: Cambridge University Press.

Thompson, J. (1990). *Ideology and modern culture.* Stanford, CA: Stanford University Press.

Thompson, P. (2000). *The voice of the past* (3rd ed.). Oxford, UK: Oxford University Press.

Thorndike, R. L. (1967). The analysis and selection of test items. In D. Jackson & S. Messick (Eds.), *Problems in human assessment* (pp. 201-216). New York: McGraw-Hill.

Thorne, S. (1994). Secondary analysis in qualitative research: Issues and implications. In J. Morse (Ed.), *Critical issues in qualitative research methods* (pp. 263-279). Thousand Oaks, CA: Sage.

Thurstone, L. L. (1925). A method of scaling psychological and educational tests. *Journal of Educational Psychology*, *16*, 433-451.

Thurstone, L. L. (1926). The scoring of individual performance. *Journal of Educational Psychology*, *17*, 446-457.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*, 273-286.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, *23*, 529-554.

Thurstone, L. L. (1929). Fechner's law and the method of equalappearing intervals. *Journal of Experimental Psychology*, *12*, 214-224.

Thurstone, L. L. (1931). Measurement of social attitudes. *Journal of Abnormal and Social Psychology*, *26*, 249-269.

Thurstone, L. L. (1947). *Multiple factor analysis.* Chicago: University of Chicago Press.

Thurstone, L. L., & Chave, E. J. (1929). *The measurement of attitudes.* Chicago: University of Chicago Press.

Timm, N. H. (1975). *Multivariate analysis with applications in education and psychology.* Monterey, CA: Brooks/Cole.

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, *20*, 410-433.

Tinsley, H. E., & Weiss, D. J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology*, *22*, 358-376.

Tintner, G. (1968). *Methodology and mathematical economics and econometrics.* Chicago: University of Chicago Press.

Titon, J. (1980). The life story. *Journal of American Folklore*, *93*(369), 276-292.

Titscher, S., Meyer, M., Wodak, R., & Vetter, R. (2000). *Methods of text and discourse analysis.* London: Sage.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, *26*, 24-36.

Tolley, H. D., & Manton, K. G. (1992). Large sample properties of estimates of discrete grade of membership model. *Annals of the Institute of Statistical Mathematics*, *41*, 85-95.

Tolley, H. D., Kovtun, M., Manton, K. G., & Akushevich, I. (2003). Statistical properties of grade of membership models for categorical data. Manuscript to be submitted to *Proceedings of National Academy of Sciences*, 2003.

Tommerup, P. (1993). *Adhocratic traditions, experience narratives and personal transformation: An ethnographic study of the organizational culture and folklore of the Evergreen State College, an innovative liberal arts college.* Unpublished doctoral dissertation, University of California, Los Angeles.

Toombs, S. K. (2001). Reflections on bodily change: The lived experience of disability. In S. K. Toombs (Ed.), *Handbook of phenomenology and medicine* (pp. 247-261). Boston: Kluwer Academic.

Toothaker, L. E. (1991). *Multiple comparisons for researchers.* Newbury Park, CA: Sage.

Toothaker, L. E. (1993). *Multiple comparison procedures.* Newbury Park, CA: Sage.

Torgerson, W. S. (1958). *Theory and methods of scaling.* New York: Wiley.

Toulmin, S. E. (1978). Science, philosophy of. In *Encyclopaedia Britannica* (15th ed., Vol. 16, pp. 375-393). Chicago: Britannica.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response.* Cambridge, UK: Cambridge University Press.

Townsend, J. T., & Ashby, F. G. (1984). Measurement scales and statistics: The misconception misconceived. *Psychological Bulletin*, *96*, 394-401.

Townsend, P., Davidson, N., & Whitehead, M. (1988). *Inequalities in health: The Black report and the health divide.* Harmondsworth: Penguin.

Tracy, K. (2002). *Everyday talk: Building and reflecting identities.* New York: Guilford.

Tracy, P. E., & Fox, J. A. (1981). The validity of sensitive measurements: A comparison of two measurement strategies. *American Sociological Review*, *46*, 187-200.

Train, K. (1986). *Qualitative choice analysis: Theory, econometrics, and an application to automobile demand.* Cambridge: MIT Press.

Tranfield, D., & Starkey, K. (1998). The nature, organization and promotion of management research: Towards policy. *British Journal of Management*, *9*, 341-353.

Traugott, M. W., & Katosh, J. P. (1979). Response validity in surveys of voting behavior. *Public Opinion Quarterly*, *43*(3), 359-377.

Trice, A. D. (1987). Informed consent: Ⅶ: Biasing of sensitive self-report data by both consent and information. *Journal of Social Behavior and Personality*, *2*, 369-374.

Triplett, N. (1898). The dynamogenic factors in pacemaking and competition. *American Journal of Psychology*, *9*, 507-533.

Troyna, B. (1994). Reforms, research and being reflexive about being reflective. In D. Halpin & B. Troyna (Eds.), *Researching education policy.* London: Falmer.

Trussell, J., & Rodríguez, G. (1990). Heterogeneity in demographic research. In J. Adams, D. A. Lam, A. I. Hermalin, & P. E. Smouse (Eds.), *Convergent issues in genetics and demography* (pp. 111-132). Oxford, UK: Oxford University Press.

Truzzi, M. (1974). *Verstehen: Subjective understanding in the social sciences.* Reading, MA: Addison-Wesley.

Tsay, R. S. (2002). *Analysis of financial time series.* New York: Wiley.

Tsiatis, A. (1975). A nonidentifiability aspect of the

problem of competing risks. *Proceedings of the National Academy of Sciences*, *72*, 20-22.

Tsuji, R. (1999). Trusting behavior in prisoner's dilemma: The effects of social networks. *Dissertation Abstracts International Section A: Humanities & Social Sciences*, *60* (5-A), 1774. (UMI No. 9929113)

Tuchman, G. (1994). Historical social science: Methodologies, methods, and meanings. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 306-323). Thousand Oaks, CA: Sage.

Tucker, L. R. (1960). Intra-individual and inter-individual multidimensionality. In H. Gulliksen & S. Messick (Eds.), *Psychometric scaling: Theory and applications* (pp. 155-167). New York: Wiley.

Tufte, E. R. (1983). *The visual display of quantitative information.* Cheshire, CT: Graphics Press.

Tukey, J. W. (1958). Bias and confidence in not quite large samples [abstract]. *Annals of Mathematical Statistics*, *29*, 614.

Tukey, J. W. (1960). A survey of sampling from contaminated normal distributions. In I. Olkin et al. (Eds.), *Contributions to probability and statistics.* Stanford, CA: Stanford University Press.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

Tuma, N. B., & Hannan, M. T. (1984). *Social dynamics: Models and methods.* Orlando, FL: Academic Press.

Turner, A. C. (1982). What subjects of survey research believe about confidentiality. In J. E. Sieber (Ed.), *The ethics of social research: Surveys and experiments* (pp. 151-165). New York: Springer-Verlag.

Turner, R. H. (1951). The quest for universals. *American Sociological Review*, *18*(6), 604-611.

Turner, S. P., & Factor, R. A. (1981). Objective possibility and adequate causality in Weber's methodological writings. *Sociological Review*, *29* (1), 5-28.

Turner, V., & Bruner, E. (Eds.). (1986). *The anthropology of experience.* Urbana: University of Illinois Press.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, *185*, 1124-1131.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, *59*, 251-278.

Tzelgov, J., & Henik, A. (1991). Suppression situations in psychological research: Definitions, implications, and applications. *Psychological Bulletin*, *109*, 524-536.

U.S. Census Bureau. (2001). *Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy on Adjustment for Non-Redistricting Uses* (With supporting documentation, Reps. 1-24). Washington, DC: Government Printing Office.

U.S. Federal Regulations of Human Research 45 CFR 46. Retrieved from http://ohrp.osophy.dhhs.gov.

UK Data Archive (n. d.). [Online]. Available: http://www.dataarchive. ac.uk/.

United Nations Statistics Division. (2001). *Handbook on census management for population and housing censuses-Series F* (No. 83, Revision 1). New York: United Nations.

United Nations. (1995). *World population prospects: The 1994 revision.* New York: Author.

United Nations. (1996). *Demographic yearbook 1994.* New York: Author.

University of Mississippi. (n.d.). PsychExperiments. Retrieved from http://psychexps.olemiss.edu.

Utts, J. M. (1996). *Seeing through statistics.* Belmont, CA: Duxbury.

Valero, P., Young, F., & Friendly, M. (2003). Visual categorical analysis in ViSta. *Computational Statistics and Data Analysis*, *43*(4), 495-508.

Vallet, L.-A. (2001). Forty years of social mobility in France. *Revue Francaise de Sociologie: An Annual English Selection*, *42*(Suppl.), 5-64.

Valsiner, J., & Benigni, L. (1986). Naturalistic research and ecological thinking in the study of child development. *Developmental Review*, *6* (3), 203-223.

Van de Geer, J. (1971). *Introduction to multivariate analysis for the social sciences.* San Francisco: Freeman.

Van den Ende, H. W. (1971). *Beschrijvende Statistiek voor Gedragswetenschappen.* Amsterdam/Brussel: Agon Elsevier.

van der Ark, L. A., & van der Heijden, P. G. M. (1998). Graphical display of latent budget analysis and latent class analysis, with special reference to correspondence analysis. In M. Greenacre & J. Blasius (Eds.), *Visualization of categorical data* (pp. 489-508). San Diego: Academic Press.

van der Ark, L. A., van der Heijden, P. G. M., & Sikkel, D. (1999). On the identifiability of the latent model. *Journal of Classification*, *16*, 117-137.

van der Heijden, P. G. M., Mooijaart, A., & de Leeuw, J. (1992). Constrained latent budget analysis. In P. Marsden (Ed.), *Sociological methodology* (pp. 279-320). Cambridge, UK: Basil Blackwell.

van der Heijden, P. G. M., van der Ark, L. A., & Mooijaart, A. (2002). Some examples of latent budget analysis and its extensions. In J. A. Hagenaars & A. McCutcheon (Eds.), *Applied latent class analysis* (pp. 107-136). Cambridge, UK: Cambridge University Press.

Van der Heijden, P. G. M., van Gils, G., Bouts, J., & Hox, J. J. (2000). A comparison of randomized response, computerassisted self-interview, and face-to-face direct questioning: Eliciting sensitive information in the context of welfare and unemployment benefit. *Sociological Methods & Research*, *28*, 505-537.

Van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory.* New York: Springer.

Van Dijk, T. A. (Ed.). (1985). *Handbook of discourse analysis: Vols. 1-4.* London: Academic Press.

Van Maanen, J. (1979). The fact and fiction in organizational ethnography. *Administrative Science Quarterly*, *24*, 539-550.

Van Maanen, J. (1988). *Tales of the field: On writing ethnography.* Chicago: University of Chicago Press.

van Manen, M. (1997). *Researching lived experience: Human science for an action sensitive pedagogy.* London, Ontario: Althouse.

van Teijlingen, E., & Hundley, V. (2002). The importance of pilot studies. *Nursing Standard*, *16* (40), 33-36.

Vapnik, V. N. (1999). *Statistical learning theory.* New York: Wiley.

Vaupel, J. W. (1990). Relatives' risks: Frailty models of life history data. *Theoretical Population Biology*, *37*, 220-234.

Vaupel, J. W., & Yashin, A. I. (1985). The deviant dynamics of death in heterogenous populations. In N. B. Tuma (Ed.), *Sociological methodology* (pp. 179-211). San Francisco: Jossey-Bass.

Vaupel, J. W., & Yashin, A. I. (1985). Heterogeneity's ruses: Some surprising effects of selection on population dynamics. *American Statistician*, *39*, 176-185.

Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, *16*, 439-454.

Velleman, P. F. (1988). *DataDesk version 6.0 statistics guide.* Ithaca, NY: Data Description, Inc.

Velleman, P. F., & Wilkinson, L. (1993). Nominal, ordinal, interval and ratio typologies are misleading. *American Statistician*, *47*, 65-72.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer-Verlag.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data.* New York: Springer.

Verbrugge, L. (1980). Health diaries. *Medical Care*, *18*, 73-95.

Vermunt, J. K. (1997). *Log-linear models for event histories.* Thousand Oaks, CA: Sage.

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89-106). Cambridge, UK: Cambridge University Press.

Vermunt, J. K., Rodrigo, M. F., & Ato-Garcia, M. (2001). Modeling joint and marginal distributions in the analysis of categorical panel data. *Sociological Methods & Research*, *30*, 170-196.

Vogt, P. W. (1993). *Dictionary of statistics and methodology: A nontechnical guide for the social sciences.* Newbury Park, CA: Sage.

Vogt, W. P. (1999). *Dictionary of statistics and methodology: A nontechnical guide for the social sciences* (2nd ed.). Thousand Oaks, CA: Sage.

Von Collani, E., & Drager, K. (2001). *Binomial distribution handbook.* Basel, Switzerland: Birkhauser.

von Glasersfeld, E. (1984). An introduction to radical constructivism. In P. Watzlawick (Ed.), *The invented reality.* New York: Norton.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behaviour.* Princeton, NJ: Princeton University Press.

von Wright, G. H. (1971). *Explanation and understanding.* Ithaca, NY: Cornell University Press.

Wachter, K. W., & Trussell, J. (1982). Estimating historical heights. *Journal of the American Statistical Association*, *77*(378), 279-293.

Wackerly, D. D., Mendelhall, W., Ⅲ, & Scheaffer, R. L. (1996). *Mathematical statistics with applications* (5th ed.). Belmont, CA: Duxbury.

Wahba, G. (1990). *Spline models for observational data.* Philadelphia: SIAM.

Walby, S., & Olsen, W. (2002, November). *The impact of women's position in the labour market on pay and implications for UK productivity.* Cabinet Office Women and Equality Unit. Retrieved from www.womenandequalityunit. gov.uk.

Wald, A., & Wolfowitz, J. (1940). On a test whether two samples are from the same population. *Annals of Mathematical Statistics*, *11*, 147-162.

Walker, E., & Lev, J. (1953). *Statistical inference.* New York: Henry Holt.

Walker, R. (Ed.). (1985). *Applied qualitative research.* Aldershot, UK: Gower.

Walkerdine, V., Lucey, H., & Melody, J. (2001). *Growing up girl: Psycho-social explorations of gender and class.* Basingstoke, UK: Palgrave.

Wallgren, A. et al. (1996). *Graphing statistics and data: Creating better data.* Thousand Oaks, CA: Sage.

Walsh, A., & Ollenburger, J. C. (2001). *Essential statistics for the social and behavioral sciences: A conceptual approach.* Upper Saddle River, NJ: Prentice Hall.

Walt, S. M. (1987). *The origins of alliances.* Ithaca, NY: Cornell University Press.

Walton Braver, M., & Braver, S. L. (1996). Statistical treatment of the Solomon four-group design: A meta-analytic approach. *Psychological Bulletin*, *104*(1), 150-154.

Walton, J. (1992). Making the theoretical case. In C. C. Ragin & H. S. Becker (Eds.), *What is a case? Exploring the foundations of social inquiry* (pp. 121-137). New York: Cambridge University Press.

Ward, A. (2002). *Oral history and copyright.* Retrieved June 7, 2002 from http://www.oralhistory.org.uk.

Warde, A., Martens, L., & Olsen, W. K. (1999). Consumption and the problem of variety: Cultural omnivorousness, social distinction and dining out. *Sociology*, *30*(1), 105-128.

Warner, R. M. (1998). *Spectral analysis of time-series data.* New York: Guilford.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, *60*, 63-69.

Warren, C. A. B. (1987). *Madwives: Schizophrenic women in the 1950s.* New Brunswick, NJ: Rutgers University Press.

Warren, C. A. B. (2002). Qualitative interviewing. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research: Context and method* (pp. 83-101). Thousand Oaks, CA: Sage.

Warren, C. A. B., & Hackney, J. K. (2000). *Gender issues in ethnography* (Qualitative Research Methods Series vol. 9, 2nd ed.). Thousand Oaks, CA: Sage.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications.* New York: Cambridge University Press.

Wasserman, S., & Galaskiewicz, J. (Eds.). (1994). *Advances in social network analysis.* Thousand Oaks, CA: Sage.

Watson, J. B. (1959). *Behaviorism.* Chicago: University of Chicago Press.

Watson, M. W. (1994). Vector autoregression and cointegration. In R. F. Engle & D. L. McFadden (Eds.), *Handbook of econometrics* (Vol. 4, pp.

2843-2915). Amsterdam: Elsevier.

Watts, D. (1999). *Small worlds: The dynamics of networks between order and randomness.* Princeton, NJ: Princeton University Press.

Weaver, L., & Cousins, B. (2001, November). *Unpacking the participatory process.* Paper presented at the annual meeting of the American Evaluation Association, St. Louis, MO.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive measures in the social sciences.* Chicago: Rand McNally.

Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, L., & Grove, J. B. (1981). *Nonreactive measures in the social sciences* (2nd ed.). Boston: Houghton Mifflin.

Webb, J. F., Khazen, R. S., Hanley, W. B., Partington, M. S., Percy, W. J. L., & Rathborn, J. C. (1973). PKU screening—is it worth it? *Canadian Medical Association Journal*, *108*, 328-329.

Weber, M. (1949). Die "Objektivität" sozialwissenschaftlicher und sozialpolitischer Erkenntnis ["Objectivity" in social science and social policy]. In *The methodology of the social sciences: Max Weber* (pp. 50-112). New York: Free Press. (Originally published in 1904)

Weber, M. (1964). *The theory of social and economic organization* (A. M. Henderson & T. Parsons, Trans.). New York: Free Press.

Weber, M. (1975). *Roscher and Knies.* New York: Free Press.

Weber, R. P. (1990). *Basic content analysis.* Newbury Park, CA: Sage.

Wedel, M., & DeSarbo, W. S. (1994). A review of recent developments in latent class regression models. In R. P. Bagozzi (Ed.), *Advanced methods of marketing research* (pp. 352-388). Cambridge, UK: Basil Blackwell.

Weisberg, H. F. (1974). Dimensionland: An excursion into spaces. *American Journal of Political Science*, *18*, 743-776.

Weisberg, H. F. (1974). Models of statistical relationship. *American Political Science Review*, *68*, 1638-1655.

Weisberg, H. F. (1992). *Central tendency and variability* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-083). Newbury Park, CA: Sage.

Weisberg, S. (1985). *Applied linear regression.* New York: Wiley.

Weiss, C. (1977). *Use of social research in public policy.* Lexington, MA: D. C. Heath.

Weiss, C. H. (1972). A treeful of owls. In C. H. Weiss (Ed.), *Evaluating action programs* (pp. 3-27). Boston: Allyn & Bacon.

Weiss, C. H., & Bucuvalas, M. (1980). Truth test and utility test: Decision makers' frame of reference for social science research. *American Sociological Review*, *45*, 302-313.

Weiss, H. M. (2002). Deconstructing job satisfaction: Separating evaluations, beliefs and affective experiences. *Human Resources Management Review*, *12*, 173-194.

Weiss, R. (1995). *Learning from strangers.* New York: Free Press.

Weitzman, E. A., & Miles, M. B. (1995). *Computer programs for qualitative data analysis: A software source book.* Thousand Oaks, CA: Sage.

Welch, B. L. (1947). The generalization of student's problem when several different population variances are involved. *Biometrika*, *34*, 28-35.

Weller, S. C., & Romney, A. K. (1988). *Systematic data collection* (Qualitative Research Methods, Vol. 10). Newbury Park, CA: Sage.

Weller, S. C., & Romney, A. K. (1990). *Metric scaling: Correspondence analysis.* Newbury Park, CA: Sage.

Wellman, B., & Berkowitz, S. D. (Eds.). (1997). *Social structures: A network approach* (updated ed.). Greenwich, CT: JAI.

Wengraf, T. (2001). *Qualitative research interviewing: Biographic-narrative and semi-structured method.* London: Sage.

Werner, O., & Schoepfle, G. M. (1987). *Systematic fieldwork* (2 vols.). Newbury Park, CA: Sage.

West, S. G., Hepworth, J. T., McCall, M. A., & Reich, J. W. (1989). An evaluation of Arizona's July 1982 drunk driving law: Effects on the City of Phoenix. *Journal of Applied Social Psychology*, *19*,

1212-1237.

Wetherell, M., & Potter, J. (1992). *Mapping the language of racism: Discourse and the legitimation of exploitation.* Brighton, UK: Harvester Wheatsheaf.

Wetherell, M., Taylor, S., & Yates, S. J. (Eds.). (2001). *Discourse theory and practice.* London: Sage.

Wheldall, K., & Merrett, F. (1989). *Positive teaching: The behavioural approach.* Birmingham, UK: Positive Products.

Wherry, R. J., Sr. (1984). *Contributions to correlational analysis.* New York: Academic Press.

Whewell, W. (1847). *The philosophy of the inductive sciences* (Vols. 1 & 2). London: Parker.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*, 817-838.

White, H. (1984). *Asymptotic theory for econometricians.* Orlando, FL: Academic Press.

White, L. (1999). *Political analysis: Technique and practice* (4th ed.). Orlando, FL: Harcourt Brace.

Whiting, J. W., Child, I. L., & Lambert, W. W. (1966). *Field guide for the study of socialization.* New York: Wiley.

Whittaker, J. (1990). *Graphical models in applied multivariate statistics.* Chichester, UK: Wiley.

Whyte, W. F. (1955). *Street corner society: The social structure of an Italian slum* (2nd ed.). Chicago: University of Chicago Press. (Original work published 1943)

Whyte, W. F. (1984). *Learning from the field: A guide from experience.* Beverly Hills, CA: Sage.

Whyte, W. F. (Ed.). (1991). *Participatory action research.* Newbury Park, CA: Sage.

Wiggershaus, R. (1995). *The Frankfurt School: Its history, theories and political significance* (M. Robertson, Trans.). Cambridge, UK: Polity.

Wiggins, L. M. (1973). *Panel analysis.* Amsterdam: Elsevier.

Wilcox, R. (2001). *Fundamentals of modern statistical methods: Substantially improving power and accuracy.* New York: Springer.

Wilcox, R. R. (1996). *Statistics for the social sciences.* San Diego, CA: Academic Press.

Wilcox, R. R. (2001). *Fundamentals of modern statistical methods.* New York: Springer.

Wilcox, R. R. (2003). *Applying contemporary statistical techniques.* San Diego, CA: Academic Press.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, *1*, 80-83.

Wildt, A. R., & Ahtola, O. T. (1978). *Analysis of covariance* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-012). Beverly Hills, CA: Sage.

Willer, D. (1967). *Scientific sociology: Theory and method.* Englewood Cliffs, NJ: Prentice Hall.

Williams, G. (1984). The genesis of chronic illness: Narrative re-construction. *Sociology of Health & Illness*, *6*(2), 175-200.

Williams, L. J., & Brown, B. K. (1994). Method variance in organizational behavior and human resources research: Effects on correlations, path coefficients, and hypothesis testing. *Organizational Behavior and Human Decision Processes*, *57*, 185-209.

Williams, M. (1998). The social world as knowable. In T. May & M. (Eds.), *Knowing the social world* (pp. 5-21). Buckingham, UK: Open University Press.

Williams, M. (2000). Interpretivism and generalisation. *Sociology*, *34*(2), 209-224.

Williams, M. (2000). *Science and social science: An introduction.* London: Routledge.

Williams, M., & May, T. (1996). *Introduction to the philosophy of social research.* London: UCL Press.

Williams, R. (1977). *Marxism and literature.* Oxford, UK: Oxford University Press.

Willson, V. L., & Putnam, R. R. (1982). A meta-analysis of pretest sensitization effects in experimental design. *American Educational Research Journal*, *19*, 249-258.

Wilson, P. J. (1974). *Oscar: An inquiry into the nature of sanity.* New York: Vintage.

Wilson, T. P. (1969). A proportional reduction in error interpretation for Kendall's tau-b. *Social Forces*, *47*, 340-342.

Winch, P. (1990). *The idea of a social science and its relation to philosophy.* London: Routledge.

(Originally published in 1958)

Winer, B. J. (1962). *Statistical principles in experimental designs*. New York: McGraw-Hill.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). New York: McGraw-Hill.

Winship, C., & Mare, R. D. (1992). Models for sample selection bias. *Annual Review of Sociology*, *18*, 327-350.

Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, *25*, 659-706.

Winship, C., & Radbill, L. (1994). Sampling weights and regression analysis. *Sociological Methods & Research*, *23*(2), 230-257.

Wiseman, F., Moriarty, M., & Schafer, M. (1975-1976). Estimating public opinion with the randomized response model. *Public Opinion Quarterly*, *39*(4), 507-513.

Wishart, J. (1938). Growth-rate determinations in nutritional studies with the Bacon pig, and their analysis. *Biometrika*, *30*, 16-28.

Wittgenstein, L. (1974). *On certainty* (G. E. M. Anscombe & G. H. Von Wright, Eds.). Oxford, UK: Blackwell.

Wodak, R., & Meyer, M. (2001). *Methods in critical discourse analysis*. Thousand Oaks, CA: Sage.

Wold, H. (1966). Estimation of principal components and related models by iterative least squares. In P. R. Krishnaiaah (Ed.), *Multivariate analysis* (pp. 391-420). New York: Academic Press.

Wolf, M. (1992). *A thrice told tale: Feminism, postmodernism and ethnographic responsibility*. Stanford, CA: Stanford University Press.

Wolfe, J. H. (1970). Pattern clustering by multivariate mixtures. *Multivariate Behavioral Research*, *5*, 329-350.

Wonnacott, T. H., & Wonnacott, R. J. (1990). *Introductory statistics for business and economics* (4th ed.). San Francisco: Jossey-Bass.

Wood, B. D. (1992). Modeling federal implementation as a system: The clean air case. *American Journal of Political Science*, *1*, 40-67.

Wood, J. W., & Weinstein, M. (1990). Heterogeneity in fecundability: The effect of fetal loss. In J. Adams, D. A. Lam, A. I. Hermalin, & P. E. Smouse (Eds.), *Convergent issues in genetics and demography* (pp. 171-188). Oxford, UK: Oxford University Press.

Wood, J. W., Holman, D. J., Yasin, A., Peterson, R. J., Weinstein, M., & Chang, M.-C. (1994). A multistate model of fecundability and sterility. *Demography*, *31*, 403-426.

Wood, S. (2001). mgcv: GAMs and generalized ridge regression for R. *R News*, *1*, 20-25.

Woodbury, M. A., & Clive, J. (1974). Clinical pure types as a fuzzy partition. *Journal of Cybernetics*, *4*(3), 111-121.

Woodward, J. (1995). Causation and explanation in econometrics. In D. Little (Ed.), *On the reliability of economic models: Essays in the philosophy of economics* (pp. 9-61). Boston: Kluwer Academic.

Wooldridge, J. M. (2000). *Introductory econometrics: A modern approach*. Cincinnati, OH: South-Western.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge: MIT Press.

Wooldridge, J. M. (2003). *Introductory econometrics: A modern approach* (2nd ed.). Cincinnati, OH: South-Western College Publishing.

Woolgar, S. (Ed.). (1988). *Knowledge and reflexivity*. London: Sage.

World Bank, International Economics Department. (1997). *World tables of economic and social indicators, 1950-1992* [Computer file]. Ann Arbor, MI: Interuniversity Consortium for Political and Social Science Research.

Wright, B. D. (1997). S. S. Stevens revisited. *Rasch Measurement Transactions*, *11*, 552-553.

Wright, S. (1918). On the nature of size factors. *Genetics*, *3*, 367-374.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, *5*, 161-215.

Wright, W. (1975). *Six guns and society*. Berkeley: University of California Press.

Wu, L. L., & Martinson, B. C. (1993). Family structure and the risk of a premarital birth. *American Sociological Review*, *58*(2), 210-232.

Wu, L. L., & Tuma, N. B. (1990). Local hazard

models. In C. C. Clogg (Ed.), *Sociological methodology 1990* (pp. 141-180). Oxford, UK: Basil Blackwell.

Wylie, R. (1974). *The self concept: A review of methodological considerations and measuring instruments.* Lincoln: University of Nebraska Press.

Xie, Y. (1989). An alternative purging method: Controlling the composition-dependent interaction in an analysis of rates. *Demography*, *26*, 711-716.

Xie, Y. (1992). The log-multiplicative layer effect model for comparing mobility tables. *American Sociological Review*, *57*, 380-395.

Xie, Y., & Pimentel, E. E. (1992). Age patterns of marital fertility: Revisingthe Coale-Trussell method. *Journal of the American Statistical Association*, *87*, 977-984.

Yates, F. (1934). The analysis of multiple classifications with unequal numbers in the different classes. *Journal of the American Statistical Association*, *29*, 51-66.

Yeo, I.-K., & Johnson, R. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, *87*, 954-959.

Yin, R. K. (1989). *Case study research: Design and methods.* Newbury Park, CA: Sage.

Yin, R. K. (1994). *Case study research: Design and methods* (2nd ed.). Thousand Oaks, CA: Sage.

Yin, R. K. (1998). The abridged version of case study research. In L. Bickman & D. J. Rog (Eds.), *Handbook of applied social research methods* (pp. 229-259). Thousand Oaks, CA: Sage.

Young, F. W., & Hamer, R. M. (1987). *Multidimensional scaling: History, theory, and applications.* Hillsdale, NJ: Lawrence Erlbaum.

Young, K., Ashby, D., Boaz, A., & Grayson, L. (2002). Social science and the evidence-based policy movement. *Social Policy and Society*, *1*(3), 215-224.

Yuen, K. K. (1974). The two-sample trimmed *t* for unequal populations variances. *Biometrika*, *61*, 165-170.

Yule, G. U. (1903). Notes on the theory of association of attributes in statistics. *Biometrika*, *2*, 121-134.

Yule, G. U. (1971). On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers. In A. Stuart & M. Kendall (Eds.), *Statistical papers of George Udny Yule* (pp. 389-420). New York: Hafner. (Original work published 1927)

Yule, G. U., & Kendall, M. G. (1950). *An introduction to the theory of statistics* (14th ed.). London: Griffin.

Yule, G. U., & Kendall, M. G. (1968). *An introduction to the theory of statistics.* New York: Hafner.

Zadeh, L. (1965). Fuzzy sets. *Information and Control*, *8*, 338-353.

Zaller, J. R. (1992). *The nature and origins of mass opinion.* Cambridge, UK: Cambridge University Press.

Zaman, A. (1996). *Statistical foundations for econometric techniques.* New York: Academic Press.

Zanna, M. P., & Cooper, J. (1974). Dissonance and the pill: An attribution approach to studying the arousal properties of dissonance. *Journal of Personality and Social Psychology*, *29*, 703-709.

Zeeman, E. C. (1972). Differential equations for the heartbeat and nerve impulse. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Vol. 4, pp. 8-67). Edinburgh, UK: Edinburgh University Press.

Zeisel, H., & Kaye, D. H. (1997). *Prove it with figures.* New York: Springer.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, *57*, 348-368.

Zellner, A., & Theil, H. (1962). Three-stage least squares: Simultaneous estimation of simultaneous equations. *Econometrica*, *30*, 54-78.

Zetenyi, T. (Ed.). (1988). *Fuzzy sets in psychology.* Amsterdam: North-Holland.

Zimmerman, D. H., & Wieder, D. (1977). The diary-interview method. *Urban Life*, *5*, 479-498.

Zimmerman, M. (2001-2002). Rigoberta Menchú, David Stoll, subaltern narrative and testimonial truth: A personal testimony. *Antípodas: Journal of Hispanic and Galician Studies*, *13/14*, 103-124.

Znaniecki, F. (1934). *The method of sociology*. New York: Farrar and Rhienhart.

Znaniecki, F. (1969). *On humanistic sociology: Selected papers* (R. Bierstadt, Ed.). Chicago: University of Chicago Press.

Zorn, C. J. W. (2001). Generalized estimating equation models for correlated data: A review with applications. *American Journal of Political Science*, *45*, 470-490.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*, 432-442.

Zwillinger, D. (1995). *CRC standard mathematical tables and formulae* (30th ed.). Boca Raton, FL: CRC Press.

索　引

汉英人名对照表*

阿博特,安德鲁　Andrew Abbott
阿卜迪,埃尔韦　Hervé Abdi
阿代尔,约翰　John G. Adair
阿德勒,彼得　Peter Adler
阿德勒,帕特里夏　Patricia A. Adler
阿迪蒂亚,拉姆　Ram N.Aditya
阿多诺,西奥多　Theodor Adorno
阿尔都塞　Althusser
阿尔金,马尔温　Marvin Alkin
阿尔特曼,迈卡　Micah Altman
阿尔特曼,斯图尔特　Stuart A. Altman
阿尔瓦雷茨,迈克尔　R. Michael Alvarez
阿格雷斯蒂,艾伦　Alan Agresti
阿加,迈克尔　Michael Agar
阿克塞尔罗德,罗伯特　Robert Axelrod
阿肯　Achen
阿拉比　Arabie
阿里　Ary
阿罗,肯尼思　Kenneth J. Arrow
阿罗诺,爱德华　Edward Aronow
阿姆斯特朗　Armstrong
阿穆勒,雅克　Jacque L. Amoureux
阿诺德　Arnold
阿什比　Ashby
阿什德,大卫　David L. Altheide
阿什莫尔,马尔科姆　Malcolm Ashmore
阿施　Asch
阿斯克吉夫斯　AskJeeves
阿苏什埃兹　Associates
阿特金森,罗伯特　Robert Atkinson
埃贝尔,罗伯特　Robert L. Ebel
埃尔德顿,帕林　Palin Elderton
埃尔克,范　L. Andries Van der Ark
埃尔斯特,乔恩　Jon Elster
埃弗龙,布拉德利　Bradley Efron
埃格,马蒂亚斯　Mattias Egger
埃朗　Erlang
埃里克森,埃里克　Erik Erikson
埃利斯,查尔斯　Charles H. Ellis
埃利斯,卡罗琳　Carolyn Ellis
埃米贝伊　Emirbayer
埃奇沃斯,弗兰西斯　Francis Edgeworth
埃塞德,菲洛梅娜　Philomena Essed
埃文斯　Evans
艾尔维森,麦茨　Mats Alvesson
艾弗森,古德蒙　Gudmund R. Iversen
艾柯,安伯托　Umberto Eco
艾肯,利昂娜　Leona S. Aiken
艾伯-埃贝斯费尔特,艾雷尼厄斯　Irenaeus Eibl-Eibesfeldt
艾利森,保罗　Paul D. Allison
艾美米亚　Amemiya
艾森哈特,凯瑟琳　Kathleen M. Eisenhardt
艾斯纳,埃利奥特　Elliot Eisner
艾耶尔,A. J.　A. J. Ayer
爱德华兹,德里克　Derek Edwards
爱德华兹,舍曼　W. Sherman Edwards
爱默生　Emerson
安德森,N. H.　N. H. Anderson
安德森,S.　S. Anderson
安德森,卡罗琳　Carolyn J. Anderson
安德森,罗伯特　Robert Andersen
安格尔,约翰　John Angle
安格罗斯诺,迈克尔　Michael V. Angrosino
安塞尔姆　Anselm
安斯沃思,玛丽　Mary Ainsworth
安索林,卢克　Luc Anselin
奥德　Ord
奥恩,马丁　Martin T. Orne
奥尔波特,高尔顿　Gordon Allport
奥尔布里顿,罗伯特　Robert B. Albritton
奥尔德里奇,艾伦　Alan Aldridge

* 按汉译的姓氏拼音排序。

奥尔森,兰德尔 Randall J. Olsen
奥尔森,温蒂 Wendy K. Olsen
奥克,利安 Ann Oakley
奥克斯 Oakes
奥克特 Orcutt
奥利金,英格拉姆 Ingram Olkin
奥尼尔,凯文 Kevin O'Neil
奥斯古德 Osgood
巴比奇,查尔斯 Charles Babbage
巴彻,约翰 Johann Bacher
巴彻尔德 Batchelder
巴顿,迈克尔 Michael Quinn Patton
巴恩斯 Barnes
巴恩特,德博拉 Deborah Barndt
巴尔,罗伯特 Robert Barr
巴尔塔吉,巴迪 Badi H.Baltagi
巴尔特,罗兰 Roland Barthes
巴尔扎克,奥诺雷 Honore de Balzac
巴克,尼古拉斯 Nicholas H. Buck
巴克,瑞安 Ryan Bakker
巴克兰 Buckland
巴克曼,罗格 Roger Bakeman
巴纳德 Barnard
巴塞洛缪 Bartholomew
巴斯卡尔,罗伊 Roy Bhaskar
巴塔查里亚,古里 Gouri K. Bhattacharyya
巴特利特 Bartlett
巴特森,丹尼尔 C.Daniel Batson
巴图内克,琼 Jean M. Bartunek
柏拉图 Plato
班克斯 Banks
邦格,马里奥 Mario Bunge
保卢斯 Paulhus
保罗,帕拉布 Pallab Paul
保陶伊,达芙妮 Daphne Patai
鲍德里亚,让 Jean Baudrillard
鲍尔比,约翰 John Bowlby
鲍尔斯,丹尼尔 Daniel Powers
鲍姆林德,黛安娜 Diana Baumrind
贝尔,苏珊 Susan Bell
贝尔德,沙尔迪 Chardie L. Baird
贝尔努利,雅各布 Jakob Bernoulli
贝尔努利,约翰 Johann Bernoulli
贝弗利,约翰 John Beverly
贝克,弗兰克 Frank Baker
贝克,纳撒尼尔 Nathaniel Beck
贝克尔,霍华德 Howard Becker
贝克尔,马克 Mark P. Becker
贝克尔,索尔 Saul Becker
贝内泽,让·保罗 Jean Paul Benzécri
贝斯,托马斯 Thomas Bayes
贝特森,格雷戈里 Gregory Bateson
本顿 Benton
本弗 Benfer
本内特 Bennett
本特勒,彼得 Peter Bentler
比克曼 Bickman
比利希,迈克尔 Michael Billig
比赛特,R.T. R. T. Bissett
彼得森,埃里克 Eric Peterson
彼得斯 Peters
毕晓普,乔治 George F. Bishop
边沁,杰里米 Jeremy Bentham
波德萨可夫 Podsakoff
波尔,弗朗克 Frank Van de Pol
波伏娃,西蒙娜 Simone de Beauvoir
波利内尔 Pollner
波林斯,布赖恩 Brian M. Pollins
波伦,布莱克 Blake D. Poland
波普尔,卡尔 Karl Popper
波森 Pawson
波斯特,艾利森 Alison E. Post
波特,乔纳森 Jonathan Potter
伯德惠斯特利,R. L. R. L. Birdwhistle
伯恩,戴维 David Byrne
伯恩鲍姆 Birnbaum
伯恩斯,文森特 Vincent Burns
伯恩斯坦,罗伯特 Robert F. Bornstein
伯尔,维维恩 Vivien Burr
伯夫,苏珊娜 Suzanna de Boef
伯格纳 Bergner
伯吉斯 Burgess
伯克,理查德 Richard A. Berk
伯克 Burke
伯克哈特,罗斯 Ross E. Burkhart
伯梅克,弗雷德里克 Frederick J. Boehmke
伯纳多 Bernardo
伯奇德 Berscheid
伯索尔,梅拉妮 Melanie Birdsall
伯特,西里尔 Cyril Burt
勃兰特 Brandt
博阿兹 Boaz
博恩施泰特 Bohrnstedt
博尔纳,乔安娜 Joanna Bornat
博加塔 Borgatta
博克,达雷尔 R. Darrell Bock

博克纳,阿瑟 Arthur P. Bochner
博克斯 Box
博克斯-斯蒂芬斯迈耶,珍妮特 Janet M. Box-Steffensmeier
博伦,肯尼斯 Kenneth A. Bollen
博曼,詹姆斯 James Bohman
博斯 Bos
博斯沃思 Bosworth
博托夫,琼 Joan L. Bottorff
布伦瑞克,埃贡 Egon Brunswik
布迪厄,皮埃尔 Pierre Bourdieu
布尔克,琳达 Linda B. Bourque
布克 Boucke
布克霍尔特 Buckholdt
布拉德伯里 H. Bradbury
布拉厄,蒂乔 Tycho Brahe
布拉斯科维,吉姆 Jim Blascovich
布莱基,诺曼 Norman Blaikie
布莱克,马克思 Max Black
布莱克,托马斯 Thomas R. Black
布莱洛克,休伯特 Hubert Blalock
布赖曼 Breiman
布兰尼克 Brannick
布朗,考特尼 Courtney Brown
布朗,拉德克利夫 A. R. Radcliffe Brown
布朗,斯蒂芬 Steven R. Brown
布朗,泰勒 Taylor Brown
布朗,威廉 William Brown
布朗 Browne
布朗芬布伦纳,尤里 Urie Bronfenbrenner
布朗利 Brownlee
布劳 Blau
布雷迪,伊万 Ivan Brady
布雷克 Bryk
布里格斯,查尔斯 Charles Briggs
布里克,迈克尔 J. Michael Brick
布里曼,艾伦 Alan Bryman
布里奇曼,珀西 Percy Bridgman
布里特 Britt
布林 Breen
布卢尔 Bloor
布卢默,赫伯特 Herbert Blumer
布鲁尔,玛丽琳 Marilynn B. Brewer
布鲁克菲尔德,斯蒂芬 Stephen D. Brookfield
布鲁纳 Bruner
布伦南,罗伯特 Robert L. Brennan
布伦特 Brent
布罗克迈尔 E. Brockmeyer
布洛斯费尔德 Blossfeld
布吕芒 Blumen
布什,乔治 George W. Bush
蔡平,蒂姆 Tim Chapin
策尔戈弗 Tzelgov
查尔顿,托尼 Tony Charlton
查尔顿,约翰 John R. H. Charlton
查马,艾伯特 Albert Chammah
陈,彼得 Peter Y. Chen
达德利 Dudley
达尔文,查尔斯 Charles R. Darwin
达朗贝尔 d'Alembert
达罗克,约翰 John N. Darroch
达马利 Damaree
达斯,约翰 John Darsee
戴 Day
戴恩 Dion
戴克,特恩 Teun Van Dijk
戴维森,詹姆斯 James Davidson
戴维斯,詹姆斯 James A. Davis
丹德里奇 Dandridge
丹尼尔 Daniell
丹尼尔斯 Daniels
道格拉斯,杰克 Jack Douglas
德赫贾 Dehejia
德克内 Deckner
德莱尼 Delaney
德莱乌 DeLeeuw
德兰蒂 Delanty
德勒兹,吉尔 Gilles Deleuze
德雷斯勒,威廉 William W. Dressler
德里达,雅克 Jacques Derrida
德鲁,保罗 Paul Drew
邓巴 Dunbar
邓津,诺曼 Norman K. Denzin
邓肯,奥迪斯 Otis Duncan
邓纳姆 Dunham
狄德罗 Diderot
狄尔泰,威廉 Wilhelm Dilthey
迪博尔德 Diebold
迪尔曼 Dillman
迪基 Dickey
迪肯 Deacon
迪克逊-伍兹,玛丽 Mary Dixon-Woods
迪拉克 Dirac
迪纳尔多 Dinardo
蒂布什容尼 Tibshirani
蒂恩,查尔斯 Charles Tien

金博尔,大卫 David C. Kimball
金德 Kinder
金蒂斯 Gintis
金赛 Kinsey
津巴多,菲利普 Phillip Zimbardo
卡曾斯,J. B. J. B. Cousins
卡茨,乔纳森 Jonathan N. Katz
卡恩 Kan
卡尔顿 Carrington
卡尔森 Carlson
卡拉托什 Lakatos
卡利斯,萨莉 Sally A. Carless
卡林,B. P. B. P. Carlin
卡林,J. B. J. B. Carlin
卡伦罗斯,乔治亚 Georgia T. Karuntzos
卡罗尔 Carroll
卡迈恩,爱德华 Edward G. Carmines
卡麦伦 Cameron
卡麦兹,凯西 Kathy Charmaz
卡明,利昂 Leon Kamin
卡明斯 Cummings
卡内曼 Kahneman
卡努哈,瓦利・卡列伊 Valli Kalei Kanuha
卡普兰,A. A. Kaplan
卡普兰,E. L. E. L. Kaplan
卡普兰,大卫 David W. Kaplan
卡乔波 Cacioppo
卡塞拉 Casella
卡斯 Kass
卡斯帕罗夫,加里 Garry Kasparov
卡斯特兰 Castellan
卡特赖特,南希 Nancy Cartwright
卡特利 Cattell
卡特维克,范 Van Katwyk
卡希尔,斯宾塞 Spencer Cahill
凯达尔,奥里特 Orit Kedar
凯恩,卡罗尔 Carole Cain
凯尔,加里 Gary Cale
凯利,乔治 George Kelly
凯斯特,史蒂芬 Stephen Koester
凯特勒,阿道夫 Adolphe Quetelet
凯西 Kashy
凯泽 Kaiser
奎因,凯文 Kevin M. Quinn
坎贝尔,唐纳德 Donald T. Campbell
康德,伊曼努尔 Immanuel Kant
考克森,安东尼 Anthony P. M. Coxon
考克斯,理查德 Richard Cox
柯尔莫哥洛夫,A. A. Kolmogorov
柯克曼 Kirkman
柯伦 Curran
柯廷 Curtin
柯西,奥古斯丁 Augustin Cauchy
科茨 Kotz
科多瓦 Cordova
科恩,敦 Jon Cohen
科恩伯格 Kornberg
科尔,米歇尔 Michael Cole
科尔宾,朱丽叶 Juliet M. Corbin
科尔蒂,路易丝 Louise Corti
科尔蒂纳,乔斯 Jose M. Cortina
科尔钦斯基,马列克 Marek Korczynski
科尔森 Colson
科尔特,詹姆斯 James E. Corter
科菲,阿曼达 Amanda Coffey
科夫通,米哈伊尔 Mikhail Kovtun
科根 Kogan
科克伦,威廉 William G. Cochran
科拉诺,威廉姆 William D. Crano
科雷尔 Correll
科利尔 Collier
柯林斯,兰德尔 Randall Collins
科马克 Cormack
科佩尔 Coppel
科普,玛莎 Martha Copp
科斯特纳 Costner
科塔兹 Cortazzi
科瓦尔恰克 Kowalchuk
克拉克,M. H. M. H. Clark
克拉克,哈罗德 Harold D. Clarke
克拉克,杰克 Jack Clark
克拉克哈特,大卫 David Krackhardt
克拉劳尔,西格弗里德 Siegfried Kracauer
克拉默,邓肯 Duncan Cramer
克拉潘扎诺,文森特 Vincent Crapanzano
克拉兹,米歇尔 Michael Clatts
克莱伯 Kleiber
克莱格,克里斯 Chris Clegg
克莱曼,史蒂芬 Stephen Clayman
克赖斯特,沙伦 Sharon L. Christ
克莱斯,甘特 Gunther Kress
克兰 Kline
克朗尼 Crowne
克劳斯,奥特姆 Autumn D. Krauss
克雷布,伊恩 Ian Craib
克雷斯韦尔 Creswell

克雷西,唐纳德　Donald Cressey
克里彭多夫　Krippendorf
克里斯蒂娃,朱丽娅　Julia Kristeva
克里斯曼　Chrisman
克利夫兰,威廉　William S. Cleveland
克林顿,乔舒亚　Joshua D. Clinton
克龙,马塞尔　Marcel A. Croon
克龙巴赫,李　Lee J. Cronbach
克龙恩伯格,彼得　Pieter M. Kroonenberg
克鲁格,理查德　Richard A. Krueger
克鲁斯卡尔,W. H.　W. H. Kruskal
克罗尔,保罗　Paul Croll
克罗尔　Croll
克罗斯比　Crosby
克洛格,克里弗德　Clifford C. Clogg
克门塔　Kmenta
克施文德,托马斯　Thomas Gschwend
克沃勒　Kvale
克泽尔曼　Keselman
肯德尔,大卫　David Kendall
肯德尔,帕特里夏　Patricia Kendall
肯尼,大卫　David. A. Kenny
肯尼迪,彼得　Peter E. Kennedy
孔多　Kondo
孔多塞　Marquis de Condorcet
库德,弗雷德里克　G. Frederick Kuder
库恩,兰德尔　Randall Kuhn
库恩,曼弗雷德　Manfred Kuhn
库恩,托马斯　Thomas Kuhn
库克,托马斯　Thomas D. Cook
库兰,铁木尔　Timur Kuran
库里万,道格拉斯　Douglas Currivan
库姆斯,克莱德　Clyde H. Coombs
库珀,哈里斯　Harris M. Cooper
库珀,米克　Mick P. Couper
库奇,卡尔　Carl Couch
库泽尔,安东　Anton J. Kuzel
奎因,W.　W. van O. Quine
拉奥,C. R.　C. R. Rao
拉波波特,阿纳托尔　Anatol Rapoport
拉博弗,威廉　William Labov
拉德维尔　Radbill
拉夫拉卡,保罗　Paul J. Lavrakas
拉夫特里　Raftery
拉戈普罗,卡林　Karin Boklund Lagopoulou
拉戈普罗斯,亚历山德罗斯　Alexandros Lagopoulos
拉亨布鲁克　Lachenbruch
拉贾拉特纳姆,纳基斯瓦瑞　Nageswari Rajaratnam
拉金,查尔斯　Charles C. Ragin
拉津　Larkin
拉康,雅克　Jacques Lacan
拉克林,哈沃德　Howard Rachlin
拉克纳格　Raghunathan
拉姆齐　Ramsey
拉普拉斯,皮埃尔　Pierre Laplace
拉塞尔,伯特兰　Bertrand Russell
拉瑟,帕蒂　Patti Lather
拉什利,卡尔　Karl Lashley
拉施　Rasch
拉斯勒,苏珊　Susanne Rässler
拉西特　Russett
拉辛斯基　Rasinski
拉扎斯菲尔德,保罗　Paul Lazarsfeld
莱道尔特,约翰尼斯　Johannes Ledolter
莱德　Layder
莱恩　Lane
莱尔德　Laird
莱考夫,乔治　George Lakoff
莱曼　Lyman
莱珀　Lepper
莱斯利　Leslie
莱文,霍华德　Howard Levene
莱文,罗伯特　Robert J. Levine
莱文,欧文　Irwin P. Levin
莱文,西奥　Theo Van Leeuwen
莱文斯坦,弗拉迪米尔　Vladimir Levenshtein
莱乌,让　Jan de Leeuw
莱因哈德,塞缪尔　Samuel Leinhardt
赖尔,吉尔伯特　Gilbert Ryle
赖克,威廉　Willhelm Reich
赖斯,哈里　Harry T. Reis
赖特,威廉　William Wright
赖特,休厄尔　Sewall Wright
赖兴巴赫　Reichenbach
赖因泽尔　Reinsel
兰伯特　Lambert
兰德,安　Ann Rand
兰登,艾尔弗　Alf Landon
兰格维,玛丽　Mary D. Langerwey
朗,卡伦　Karen Long
朗,斯科特　J. Scott Long
朗格海涅,罗尔夫　Rolf Langeheine
朗热利那,克瑞斯汀　Kristin M. Langellier
劳安,范　Van Loan
劳丹　Laudan
劳登布什　Raudenbush

劳顿,德波拉 C. Deborah Laughton
劳弗 Laufer
劳克斯 Lax
劳里,斯特芬 Steffen L. Lauritzen
劳里,希瑟 Heather Laurie
劳里斯 Lawlis
劳里奥拉,马尔科 Marco Lauriola
劳希 Lawshe
勒夫,唐纳 Donna Luff
勒劳依 Leroy
勒特雷尔,温迪 Wendy Luttrell
雷德菲尔德,罗伯特 Robert Redfield
雷德罗斯科,大卫 David P. Redlawsk
雷克斯,约翰 John Rex
李,雷蒙德 Raymond M. Lee
李,理查德 Richard Lee
李克特 Likert
李振 Zhen Li
里德 Read
里格登 Rigdon
里克尔,保罗 Paul Ricoeur
里莱,吉尔伯特 Gilbert Riles
里普利 Ripley
里普斯 Rips
里斯,雅各布 Jacob Riis
里斯曼,凯瑟琳 Catherine Kohler Riessman
理查森,劳雷尔 Laurel Richardson
理查森,马里奥 Marion Richardson
利奥塔,让·弗朗索瓦 Jean François Lyotard
利布利希 Lieblich
利克斯 Lix
利默 Leamer
利普希茨 Lipsitz
利特尔,丹尼尔 Daniel Little
利特尔,贾妮 Jani S. Little
利希,埃琳 Erin Leahey
连斯基,格哈德 Gerhard Lenski
梁 Liang
廖福挺 Tim Futing Liao
列维纳斯,伊曼努尔 Emmanuel Levinas
列维-斯特劳斯,克劳德 Claude Lévi-Strauss
林达尔 Lindahl
林德,詹姆斯 James Lind
林德尔 Lindell
林德奎斯特 E. F. Lindquist
林德利 Lindley
林德史密斯,艾尔弗雷德 Alfred Lindesmith
林登,范 Van der Linden
林恩,彼得 Peter Lynn
林格斯,詹姆斯 James C. Lingoes
林吉斯,阿洛哈纳 Alphonso Lingis
林肯,伊冯娜 Yvonna S. Lincoln
林奇,迈克尔 Michael Lynch
林知己夫 Chikio Hayashi
刘从(音译) Cong Liu
刘易斯,C. I. C. I. Lewis
刘易斯-伯克,米歇尔 Michael S. Lewis-Beck
刘易斯,大卫 David Lewis
龙斯克,丹尼斯 Dennis W. Roncek
卢尼 Looney
卢苏 Rousseeuw
卢因,库尔特 Kurt Lewin
鲁宾 Rubin
鲁宾逊,W. P. W. P. Robinson
鲁宾逊,W. S. W. S. Robinson
鲁宾逊,道恩 Dawn T. Robinson
鲁道什,陶马什 Tamás Rudas
鲁斯基 Ruskey
路 Lu
路德,马丁 Martin Luther
路易斯,M. R. M. R. Louis
路易斯,T. A. T. A. Louis
路易斯,安德烈·瓦莱 Louis André Vallet
伦德贝格,乔治 George Lundberg
伦宁格,莉安娜 LeeAnn Renninger
罗宾斯 Robins
罗伯特 Robert
罗布森 Robson
罗德里格斯,格尔曼 Germán Rodríguez
罗蒂,R. R. Rorty
罗杰克,克里斯 Chris Rojek
罗杰斯 Rodgers
罗勒 Rohrer
罗姆尼 Romney
罗奇福德 Rochford
罗森 Rosen
罗森堡,米尔顿 Milton J. Rosenberg
罗森堡,莫里斯 Morris Rosenberg
罗森鲍姆,保罗 Paul Rosenbaum
罗森伯格 Rosenberger
罗森茨魏希,索尔 Saul Rosenzweig
罗森汉恩 Rosenhan
罗森塔尔,罗伯特 Robert Rosenthal
罗思曼 Rothman
罗思坦,汉娜 Hannah R. Rothstein
罗斯福,富兰克林 Franklin D. Roosevelt

米德，玛格丽特 Margaret Mead
米德，乔治 George Herbert Mead
密尔，约翰 John Stuart Mill
米尔格拉，斯坦利 Stanley Milgram
米尔斯，赖特 C. Wright Mills
米尔沃德，琳内 Lynne Millward
米勒，D. D. Miller
米勒，R. G. R. G. Miller
米勒，奥图尔 Arthur G. Miller
米勒，丹尼尔 Daniel J. Mueller
米勒，拉尔夫 Ralph O. Mueller
米勒，马克 Marc Millar
米利尔 Mulier
米利肯 Milliken
米利特，凯特 Kate Millett
米切尔 Mitchell
米切纳 Michener
米舍 Mische
米什勒，埃利奥特 Elliot Mishler
米特罗夫 Mitroff
明瑟 Mincer
摩根 Morgan
摩根斯顿，奥斯卡 Oskar Morgenstern
莫尔 Mohr
莫尔斯，贾尼丝 Janice M. Morse
莫菲特，肯尼思 Kenneth W. Moffett
莫肯 Mokken
莫勒纳尔 Molenaar
莫雷诺，埃里卡 Erika Moreno
莫雷诺，雅各布 Jacob L. Moreno
莫里斯，马丁娜 Martina Morris
莫利，大卫 David Morley
莫内特，乔治 Georges Monette
莫斯特勒，弗雷德里克 Frederick Mosteller
莫伊，莱缪尔 Lemuel A. Moyé
默里，查尔斯 Charles Murray
默滕斯，唐纳 Donna M. Mertens
木滕，本特 Bengt Muthén
穆德，亚历山大 Alexander Mood
缪尔曼 Meulman
穆尔曼 Moorman
穆霍帕德伊，考斯基 Kausiki Mukhopadhyay
穆莱克，斯坦利 Stanley A. Mulaik
穆尼，克里斯托弗 Christopher Z. Mooney
穆斯塔卡，克拉克 Clark Moustakas
穆伊亚特 Mooijaart
纳尔逊 Nelson
纳格勒，乔纳森 Jonathan Nagler
纳纳，哈林德尔 Harinder Nanda
纳森，劳拉 Laura Nathan
纳什，克里斯 Chris Nash
奈塞尔 Neisser
南布迪里 Namboodiri
内尔德 Nelder
内曼，耶日 Jerzy Neyman
尼采，弗里德里克希 Friedrich Nietzsche
尼科尔森，L. L. Nicholson
尼什撒多，希祖海科 Shizuhiko Nishisato
牛顿，R. R. R. R. Newton
牛顿，艾萨克 I. Newton
纽比，哈沃德 Howard Newby
纽博尔德 Newbold
纽拉特，奥图 Otto Neurath
农纳利 Nunnally
努佐 Nuzzo
诺布利特，乔治 George W. Noblit
诺顿 Norton
诺尔波特，赫尔穆特 Helmut Norporth
诺克，大卫 David H. Knoke
诺曼 Norman
诺特 Knott
诺维克 Novick
诺伊费尔德 Neufeld
诺伊莫，安德鲁 Andrew Noymer
欧拉，莱昂哈德 Leonhard Euler
帕蒂森，菲莉帕 Philippa Pattison
帕尔默，哈维 Harvey D. Palmer
帕尔达姆，马丁 Martin Paldam
帕夫利契夫，阿列克谢 Alexei Pavlichev
帕金，弗兰克 Frank Parkin
帕金森 Parkinson
帕克，罗伯特 Robert Park
帕克，伊恩 Ian Parker
帕克斯顿，帕梅拉 Pamela Paxton
帕里，肯 Ken W. Parry
帕卢克，伊丽莎白 Elizabeth Levy Paluck
帕吕埃尔，弗朗索瓦 Francois Cretté de Palluel
帕热斯 Pagès
帕森霍 Sung Ho Park
帕森斯，塔尔科特 Talcott Parsons
帕斯卡，布莱士 Blaise Pascal
帕斯洛，奥西斯 Osiris Jorge Parcero
帕特里克 Patrick
帕特南，H. H. Putnam
派克，肯尼思 Kenneth L. Pike
派伊，保罗 Paul E. Pye

史密斯,彼得 Peter W. F. Smith
史密斯,克里斯汀 Christine Smithies
史密斯,马克·伊斯特比 Mark Easterby Smith
史密斯,玛丽 Mary L. Smith
史密斯,约翰 John K. Smith
士吉 Suci
舒茨,阿尔弗雷德 Alfred Schütz
舒曼 Schuman
舒晓玲 Xiaoling Shu
思奈德斯,汤姆 Tom A. B. Snijders
斯蒂尔曼,托德 Todd Stillman
斯尔芭,弗兰克 Frank Srba
斯金纳 Skinner
斯科特,约翰 John Scott
斯克洛多夫斯卡,埃尔兹别塔 Elzbieta Sklodowska
斯莱格特,特雷西 Tracy Hoffmann Slagter
斯里文,米歇尔 Michael Scriven
斯梅尔 Smale
斯米尔诺夫,N. V. N. V. Smirnov
斯内德克,乔治 George W. Snedecor
斯尼思 Sneath
斯诺 Snow
斯帕诺斯 Spanos
斯宾塞,赫伯特 Herbert Spencer
斯佩克特,保罗 Paul E. Spector
斯皮德,特里 Terry P. Speed
斯皮尔曼,查尔斯 Charles Spearman
斯普拉德利 Spradley
斯塔克,菲利普 Philip B. Stark
斯坦 Stein
斯坦丁,莱昂内尔 Lionel G. Standing
斯坦利,利兹 Liz Stanley
斯坦利,朱利安 Julian C. Stanley
斯特劳斯,安塞尔姆 Anselm L. Strauss
斯特劳斯,大卫 David Strauss
斯特罗门兹,大卫 David B. Strohmetz
斯通 Stone
斯图尔特,伊丽莎白 Elizabeth A. Stuart
斯图尔特 Stewart
斯图弗尔 Stouffer
斯托尔,大卫 David Stoll
斯托克 Stork
斯托克斯,琳内 S. Lynne Stokes
斯托拉德 Stallard
斯托维尔,凯瑟琳 Katherine Stovel
斯旺森 Swanson
斯威尼,凯文 Kevin J. Sweeney
斯韦德贝里,理查德 Richard Swedberg
斯维斯古德,格雷 C. Gray Swicegood
苏恩 Suen
苏利耶 Soulliere
索贝尔 Sobel
索恩,萨莉 Sally E. Thorne
索弗罗尼奥 Sofroniou
索卡 Sokal
索伦森,杰斯珀 Jesper B. Sørensen
索绪尔,费尔迪南 Ferdinand de Saussure
塔卡宁 Takane
塔克,雅克 Jacques Tacq
塔克 Tucker
塔姆哈尼 Tamhane
塔奇曼,加夫 Gaye Tuchman
塔沙可利 Tashakkori
塔伊曼 Tayman
泰勒,查尔斯 Charles Taylor
泰斯卢克 Tesluk
泰伊灵恩,埃德温 Edwin R. van Teijlingen
坦嫩鲍姆 Tannenbaum
汤姆 Thom
汤普森,保罗 Paul Thompson
汤普森,布鲁斯 Bruce Thompson
汤普森,约翰 John Thompson
特德利 Teddlie
特恩诺 Therneau
特拉克塞尔,尼科勒 Nicole M. Traxel
特拉塞尔 Trussell
特赖恩 Tryon
特赖斯 Trice
特劳戈特,迈克尔 Michael W. Traugott
特雷西,卡伦 Karen Tracy
特里普利特 Triplett
特里维迪 Trivedi
特默拉普,彼得 Peter Tommerup
特嫩豪斯 Tenenhaus
特威蒂,理查德 Richard Tweedie
特沃斯基,A. A. Tversky
廷伯根 Tinbergen
廷斯利 Tinsley
图尔明 Toulmin
图尔纳,拉尔夫 Ralph Turner
图基,约翰 John W. Tukey
图朗若 Tourangeau
图里西 Turrisi
图马,南希 Nancy Brandon Tuma
图塞克,拉里 Larry E. Toothaker
涂尔干,埃米尔 Emile Durkheim

萧 Hsiao
谢伯克,托马斯 Thomas A. Sebeok
谢菲,亨利 Henry Scheffé
谢菲,威廉 William Shaffir
谢弗,J. L. J. L. Schafer
谢弗,威廉 William D. Schafer
谢格罗夫,伊曼纽尔 Emanuel Schegloff
谢林,托马斯 Thomas Schelling
谢林,香农·豪尔 Shannon Howle Schelin
谢泼德,罗格 Roger N. Shepard
谢宇 Yu Xie
辛格,B. B. Singer
辛格,E. E. Singer
休伯特,拉里 Larry Hubert
希韦尔,威廉 William H. Sewell
休厄尔,威廉 William Whewell
休姆,大卫 David Hume
休斯,J. A. J. A. Hughes
休斯,K. K. Hughes
休斯,埃弗里特 Everett C. Hughes
休斯,里迪安 Rhidian Hughes
雅格布森,罗曼 Roman Jakobson
雅各比,威廉 William G. Jacoby
雅各布,哈丽雅特 Harriet A. Jacob
雅卡尔,詹姆斯 James J. Jaccard
雅可比 Jacobian
雅曼格奇,卡佐 Kazuo Yamaguchi
亚伯拉罕 Abraham
亚当森 Adamson
亚里士多德 Aristotle
延加 Iyengar
扬,F. W. F. W. Young
扬,K. K. Young
扬克,罗尔夫 Rolf Janke
姚肖,吉列米娜 Guillermina Jasso
耶尔,乌迪 G. Udny Yule
耶尔姆斯勒夫,路易斯 Louis Hjelmslev
叶芝 Yates
希尔 Hill
伊尔桑 Hilsum
伊莱亚森,斯科特 Scott R. Eliason
伊利格瑞,露西 Luce Irigaray
伊曼纽尔 Emanuel
殷 Yin
因伯 Imber
元 Yuen
约,斯蒂芬 Stephen Yeo
约翰森,苏林 Søren Johansen
约翰斯顿 Johnston
约翰逊,D. A. D. A. Johnson
约翰逊,N. L. N. L. Johnson
约翰逊,保罗 Paul E. Johnson
约翰逊,杰弗里 Jeffrey C. Johnson
约翰逊,理查德 Richard A. Johnson
约翰逊,马克 Mark Johnson
约翰逊,塞缪尔 Samuel Johnson
约雷斯括格,卡尔 Karl Jöreskog
赞纳 Zanna
泽德克 Zedeck
泽尔纳 Zellner
泽沃纳 Zavoina
扎德,卢特菲 Lotfi Zadeh
詹金斯,理查德 Richard Jenkins
詹金斯 Jenkins
詹姆森 Jameson
詹姆斯,威廉 William James
詹森 Jensen
昭 Tiao
珍妮 Jenny
兹纳涅茨基,弗洛里安 Florian Znaniecki
祖佩,帕特里克 Patrick Suppes
左拉,埃米尔 Emile Zola
佐恩,克里斯托弗 Christopher Zorn
佐恩 Zorn

译后记

科学之为激情，方法之为技能

历时六年的努力，本书终于译出与读者见面。作为本书的译者之一，我在阅读和翻译本书的过程中受益匪浅。若自配第的《政治算术》(1672 年)出版算起，社会科学方法的发展已有三百多年的历史，至今已蔚然大观。定性的、统计的、实验的、模拟的诸多方法，以及与之相关的方法论和认识论共同推动着社会科学事业的发展。个人在这方法的洪川巨流面前显得如此渺小。我们很可能学有所博，但极少能做到学有专精。然而，社会科学研究又常常要求我们具有方法自觉意识，否则我们的知识发现就会沦落为预言家的呓语或鼓动者的意识形态。因此，作为社会科学的研究者，我们需要经常问自己，什么是科学？什么是社会科学？社会科学与它的研究方法之间有怎样的关系？每个社会科学的研究者都可能有自己的答案，下面是我的回答，抛砖引玉，将它作为本书后记，就教于诸位读者，以期推动社会研究方法的学习和应用。

一、科学是激情

现代科学兴起于十六世纪的西方世界，而后由于十七世纪天才科学家们的贡献，便空前地繁荣起来，不仅彻底地改变和更新了人类对其周围世界的知识，还彻底改变了人类自身的生活进程。与人对艺术、宗教、哲学等美好事物的追求一样，现代科学发展的首要动力是人们的求知激情，即对发现事物一般原理的持久而坚韧的渴望。激情是一种连接和改变人同事物之间关系的情绪。在这种情绪的影响下，事物或者变得富有吸引力，或者变得令人厌恶。科学家的激情是求知激情①。由于求知激情的持续作用，科学家的目光紧紧盯着吸引他的事物，企图从中发现事物背后的奥秘。没有求知激情的鼓动，科学家不可能心无旁骛地专注于他欲从中有所发现的事物；没有求知激情的鼓舞，科学家不可能去学习、熟悉、掌握和发明用于发现事物奥秘的工具；没有求知激情的推动，科学家不可能将具有科学价值的事实与不具有科学价值的事实区别开来；没有求知激情的坚韧，科学家也不可能忍受孤独和寂寞而反复地观察和思索手边的数据；同样，没有求知激情就不可能有对已发现的美的赏析，不可能有参与批评和讨论的热情，不可能有获得发现后的欢欣雀跃。按照迈克尔·波兰尼的说法，求知激情是科学发现的逻辑基础②。

---

① 怀特海在《理性之功用》一书中称其为“欲之欲(the appetition of appetitions)”。WHITEHEAD A N. *The Function of Reason*[M]. Boston:Beacon Press,1971:33.

② 参见迈克尔·波兰尼《个人知识》第六章“求知热情”。迈克尔·波兰尼.个人知识[M].许泽民,译.贵阳:贵州人民出版社,2000.

科学激情促使科学家反复地考察"不可约且顽固的事实(irreducible and stubborn facts)"①。尊重"不可约且顽固的事实"不仅是所有时代成功实践者的基本特质,更是所有大科学家的根本气质。爱因斯坦在给他的朋友 M.贝索的一封信中写道:"一个希望受到应有信任的理论,必须建立在有普遍意义的事实之上。"科学所追寻的事实主要来自科学家耐心、细致、系统地观测和经验积累。据说伽利略年轻时曾坐在比萨大教堂里,用自己的脉动计算从屋顶垂下的带有长链的灯的摆动时间,由此发现不管链的长短如何,每次摆动所花时间似乎相同。1576 年,天文学家第谷用丹麦国王弗里德里希二世的赏赐,在赫威恩岛上建造了城堡和天文台并一直从事天文观测,直至 1597 年。类似的例子不胜枚举。

科学家的求知激情还受到理性的引导。我们可以基于以下若干理由说,科学人是理性人。如果我们避开关于理性的种种纷繁复杂的讨论,就可以简单地辨识出理性的若干关键要素。第一个要素,科学家确信事物之中存在着一定的秩序。同时,正是这一秩序支配着事物的重复显现和变化。第二个要素,科学家确信支配事物的普遍原则可以用人类发明的符号体系予以清楚的描述和传播。这就是说支配事物的普遍原则是可知的。第三个要素,逻辑和数学是理性的基石。无论是从经验事实到一般原理的归纳,还是从一般假设到具体事实的演绎,都必须符合逻辑,人们对此应当是不存在异议的。数学使得科学计算成为可能,因此要对事物的未来发展作出可靠的预测,如果没有数学的帮助也无异于是海市蜃楼②。第四个要素,雅与美。很难对科学理性中雅与美作出明确的阐释,但我们似乎可以引用爱因斯坦的有力表述来领会它们的作用,即"人们总想以最适当的方式来画出一幅简化的和易领悟的世界图像……"③。

科学的求知激情在理性的引导下,通过系统地收集和考察经验事实,所作出的发现通常称为阶段性"理论"。理论通常由两部分构成:概念和模型。所谓概念是对一组属性相当的经验事实的抽象。当然,作为理论构造物的概念绝不是孤立的,而是组成一套互相关联,并能结合其使用的分析工具。模型是以一定的假设为前提,使用变量和参数的语言构造的一些简化的和易于理解的图示,它可以用来描述现实的某些方面。概念反映了我们对所研究事物的洞见。我们对事件变化的预测能力则来自模型。两者构成了阶段性理论的既有区别又互相补充的部分。

二、不自欺的社会科学

顺着上面的观点,我们认为,社会科学④是一门科学。这门科学研究的是社会中的人和社会。它的研究对象、研究方法和研究成果可能与自然科学的对象、方法和成果间有着巨大

---

① 参见 WHITEHEAD A N. Science and the Modern World[M]. New York: The New American Library of World Literature, 1997:3.

② 冯·诺依曼认为科学的任务在于用模型来描绘观测到的现象,而模型是辅之以语言文字解释的数学构造。见 von NEUMANN J. Method in the Physical Sciences[D]. //Theory of games, astrophysics, hydrodynamics and meteorology. London: Pergamon press, 1963:492.

③ (德)爱因斯坦.爱因斯坦文集[M].许良英,范岱年,译.北京:商务印书馆,1976:101.

④ 依照《社会科学国际大百科》的词条,社会科学涵盖了多个领域,主要包括社会学、人类学、政治科学、心理学和经济学,但通常不把历史和语言学算作社会科学。见 Kerr, Keith. Social Science, In DARITY, W. A. (eds). International encyclopedia of the social sciences[M]. 2nd ed. Detroit: Macmillan Reference USA, 2008.

的差别。由于社会科学的成果缺乏自然科学定律所能达到的精确性,它通常被称为非精确科学或软科学。但是,我们不认为社会科学在求知的意图、求知的努力以及求知的精神气质上与自然科学之间应有什么区别。

然而,我们认为社会科学同人文学科应当有所区别。这两者间的区别是重要的,因为从事社会科学的研究者常常忽视科学要求和人文关怀之间的界限。无可否认人文学科(包括语言、文学、哲学等)构成了所有科学研究者的素养中一个很大的部分,对社会科学研究者来说尤其如此。以爱因斯坦的相对论或量子力学的发展为例。虽然我们无法确知哲学在这些发现中所起的分量到底有多大,但是我们可以肯定哲学是扮演了重要角色的。通过阅读哲学著作和进行哲学思考,这无疑可以培养我们质疑的精神和思辨的技艺,扩展我们的心灵范围,让我们的目光投向遥远的星空,而不是整日迷乱于日常琐事之中。文学对科学研究者,特别是社会科学研究者也很重要。我们无法否认一位好的社会科学家需要深入到文学之中,在伟大的文学中感受人性的变幻莫测和社会的纷繁复杂。但是,我们仍然认为不能用哲学思辨和人文关怀来替代科学的要求。对此的理由可以有很多。在此,我们仅指出一点,即社会科学必须具有现实感,反面来说社会科学必须不自欺。现代社会科学的创始人之一马基雅维利在其著作《论李维》中反复强调了不自欺的极端重要性。马基雅维利在谈到自己的书时说:"请相信,唯一让我感到满意的是,我知道自己在许多场合多有自欺,在此事上我却未出差错……"①。在第二卷的前言中,他又说:"世人历来厚古薄今,虽然他们并非总有道理;他们偏爱旧事物的方式,使他们不但赞美作家的记载使他们得以知晓的时代,而且赞美步入暮年后回忆起的青春时光。他们这种看法在多数时候都是错误的,然而我相信,导致他们自欺的原因不一而足。在我看来,首要原因在于对古代事物的不理解。"②在当前的社会科学研究中,自欺可能表现为多种形式,既有对过去的无限向往,也有对未来的美好憧憬。然而,无论哪种形式的自欺,在我们看来,过多地将哲学思辨和人文关怀带入我们的理论成果之中是主因。从事社会科学研究的人常常被某些主义和情怀所左右,不自觉地将他们对美好事物的向往带入自己的理论之中。我们不否认对一种好的社会状态的追求是社会科学研究者的主要动机,但这不等于说我们可以把"向往"本身当作理论的前提或结论。借用米塞斯的话,"科学绝不告诉人应该如何行为;它只指出如果你想达到某一既定目的,你就得如何行为。"③

三、方法是技能

那么方法在我们对社会科学的追求中起什么作用呢?这正是本文想要回答的关键问题。英国大科学家皮尔逊给出的答案是,"整个科学的统一仅在于它的方法,不在于它的材料"④。这一说法可能有过度强调方法之嫌。我们更愿意接受迈克尔·波兰尼的说法,"科学靠科学家的技能来操作。科学家正是通过行使自己的技能而造就了自己的科学知识"⑤。

---

① (意)马基雅维里.论李维[M].冯克利,译.北京:世纪出版集团,2005:41

② 同上,p205.

③ 米塞斯.人的行为[M].夏道平,译.台北:远流出版事业股份有限公司,1997:50.

④ 皮尔逊.科学的规范[M].李醒民,译.北京:商务印书馆,2015:13.

⑤ 迈克尔·波兰尼.个人知识[M].许泽民,译.贵阳:贵州人民出版社,2000:49.

什么是技能？技能就是掌握了的方法。更确切的说，技能是一门本领的规范操作在“我”身上的体现。技能始终是“我”的技能，是“我”通过向老师（权威）学习，通过向范例学习，通过不断地实践，通过改正错误，通过不断地琢磨和领悟才变成“我”的实践知识的东西。波兰尼给出的两个互补例子可用以说明技能的重要性。一个例子是，他本人曾在匈牙利见到过一台崭新的、吹制电灯泡的进口机器。同一种机器在德国已经成功地运行了多年，而在匈牙利，运转了一年后仍无法生产出一只没有瑕疵的灯泡。这正是匈牙利工人缺乏机器的使用技能的例子。波兰尼说的另一个例子是，“看着今人以漫无休止的努力用显微镜学和化学、用数学和电子学仿制着清一色的与二百年前那位半文盲的斯特拉迪瓦里作为日常工作制作出来的相似的小提琴，这情景真有点使人伤感”①。这突出显示了技能是把我们和某个传统联系起来的东西。方法是外在的操作规范，技能是内化成“我”的一部分的方法。没有从事社会科学的技能就谈不上从事科学研究和得到科学成果，甚至谈不上求知激情。不会下棋的人，你让他如何保持对棋的热情呢？这样看起来掌握各项从事研究的本领就极为关键了。到此，我们已经把想要陈述的观点表述出来了，即从事社会科学研究的人要掌握各种同研究相关的技能。这个论点看起来那么微不足道！如果是这样，就请让我们用一些具体的例子来说明这有多么重要吧！这些例子都与本书的词条有关，有兴趣的读者可以参看相应词条。

1.总体研究设计

抽样调查（survey）是用问卷的形式从调查对象全体（总体）的一个或若干个样本中收集社会数据的方法。从科学研究的角度讲，我们希望获得的数据既充分（与一个研究主题相关的概念是否都测量到了）又精确。从调查实践的角度讲，我们又受到投入的经费、时间、智力资源，以及政治和社会环境等诸多因素限制。如何平衡科学研究的要求和调查实践的限制呢？一个可行的方法是将影响数据充分性和精确性的各个要素和调查的各个阶段分解开，明确它们各自对调查误差的影响，在调查设计之初就综合考虑各要素和各阶段的平衡，从而得到一个能够满足科学研究要求的“适当的”数据。这就是总体研究设计的思想。总体研究设计的思想要求我们对调查的设计、问卷的设计、抽样工作、调查实施、调查员培训和调查队伍建设以及数据的编码、录入和清洁做通盘的考虑。调查者首先需要做的不是控制调查个别阶段可能产生的误差，而是调查各个阶段可能产生的总误差。要控制调查的总误差，就需要调查者全面掌握同调查有关的各种实际技能。

有些技能可能作为方法在书本上有介绍，有些纯粹是实践性的（比如组织研究的技能和组织调查实施的技能）。作为方法，它们叙述起来可能很简单，但我们要把它们变为自己的技能就不仅要跟着专家学习，还要不断实践。我们还要看到，调查所涉及的技能众多，几乎不可能由一个人完全掌握，所以调查注定是一项合作的事业。在调查中，我们特别要学会组织和合作的技能。

2.详析

详析（也称拉扎斯菲尔德详析法）不是一种具体的统计技术，而是一个一般的探究变量间关系的推理方法。

这种方法最简单的形式是：用一个两分的检验因子（test factor）来考察两个具有时间顺

---

① 迈克尔·波兰尼.个人知识[M].许泽民，译.贵阳：贵州人民出版社，2000：79.

序的两分变量 X 和 Y 之间的关系。简单地说,就是用一个两分变量把一个二维列联表拆分成两个新的二维列联表,再进一步考察变量 X 和变量 Y 之间关系的变化,以此来厘清变量 X 和 Y 之间的关系。例如,在本书提供的例子中,变量 X 是性别,变量 Y 为是否发生车祸,检验因子是开车的里程数(高和低)。当我们单独看性别和车祸发生的列联表时,发现女性比男性更少发生车祸,似乎是更好的司机。但是,当我们用驾驶里程数作为检验因子,分别考察高里程时性别和车祸的列联表,还有低里程时性别和车祸的列联表,就会发现性别和车祸的关联消失了。原来性别和车祸的关系被"解释"为性别同驾驶里程数的关系,以及驾驶里程数同车祸的关系。用拉氏原来的表述,即任何两个变量间的关系都可以用第三个检验因子分解(两分的)作两个偏关系和两个边缘关系:$[xy]=[xy;t]'\times[xy;t]''\times[xt]\times[ty]$。

这种思考方式似乎很简单,一点也不新鲜,几乎人人都会涉及。但问题是,如何把偶然的思考方式变成一般的、系统的和常规的统计分析推论方式。拉氏详析法的贡献即在于此。拉氏一般地、系统地考察了检验因子的引入和引入后的结果。他根据检验因子居于变量 X 和 Y 之间的时序位置,把检验因子分作两类,居前的(A)和居间的(I);他又根据引入检验因子后,分解得到的关系种类,分别称作偏关系(P),即引入检验因子后,原来两变量关系不消失,和边缘关系(M),即引入后,原来两变量关系消失,产生两个新的变量关系$[xt]$和$[ty]$。如此一来,他得到了四种详析类型,分别是:(1)PA 型,引入居前检验因子后,两变量关系依然存在,但是可能有所区别。比如,在考察不同类型的电影(刺激)对人的宣传效应(反应)时,可用受教育程度作居前检验因子,看看在不同条件下,效应发生的具体方式的区别。(2)PI 型,引入居间检验因子后,原来关系仍存在,但发生分化。比如,默顿在考察职业地位同社区生活参与程度的关系时发现,不满意自己工作的白领会更多参与到社区生活中,工人则是当他们满意自己的工作时才会更多参与社区生活。这里工作满意度是居间检验因子。(3)MA 型,引入居前的检验因子后,原来变量的关系消失。这时人们可以说原来的关系是虚假关系(spurious relationship)。例如,一场火灾引来的救火车越多,火灾就越危险。这里的居前检验因子可以是火灾的大小。火灾越大引来的救火车越多,火灾越大也越危险。救火车的多少同火灾的危险程度间不存在因果关系。(4)MI 型,引入居间检验因子后,原来的关系消失。这就是本书词条的例子,性别同车祸没关系,而里程长短同车祸有关系。后两种类型都会产生新的关系,即检验因子同因变量 y 的关系。拉氏告诉我们,作为一位好的科学工作者,我们应当立即用相同的程序来考察新的关系。

利用拉氏详析法,我们还可以判断两个变量的因果关系(causal relationship)。按照拉氏的界定,即如果两个具有时序的变量 X 和 Y 之间存在相关关系,当引入任意居前检验因子后,X 和 Y 之间的关系不消失,那么可以称 X 和 Y 之间有因果关系。然而,更容易实现的是,引入一个居前检验因子来证明 X 和 Y 的关系是虚假关系。

用这个例子,我们想说明定量分析不是简单的模型设定、数据处理、结果输出和解释结果。数据分析首先是系统地考察变量间的逻辑关系,是一种推理的技艺。好的数据分析需要"清晰的"推理。好的科学工作者则需要通过学习和锻炼来造就自己的"清晰头脑"。

3.数据分析

定量分析软件和个人计算机的发展使得许多原来令人望而生畏的统计分析技术变成简单的软件操作步骤。似乎只要有数据和有软件,再加上一台计算机通过点击鼠标就可完成

数据分析。其实则不然,一个高质量的分析是一个十分艰辛的过程,既需要分析者深刻理解自己应用的分析方法的统计原理,也需要有逻辑清晰的理论假设,在各种可能的分析方法中进行比较取舍,对模型不断进行修正改进,此外也需要长期的实践。否则,所谓的分析就会变成所谓的垃圾进、垃圾出。

粗略地说,一个合理且有意义的回归分析大致包括以下四个部分。

第一,我们需要为一个理论问题找到一组适当的数据予以验证,以加强我们对理论问题的理解和信心。这涉及如何为相关的理论概念找到一组适当的测量指标,用测量指标间的关系来表示理论概念间的关系。但是,很多情况下,我们很难为理论概念找到直接的测量指标,这时就需要用别的指标来代替。于是,我们就必须理解替代指标的性质,以及它同其他指标之间的关系。此外,我们可能还需要明确指标变量间的关系,它们是线性的,还是非线性的;它们是递增的,还是递减的;一个指标的变化随着另一个指标变化的大致范围,特别是有理论意义的范围等。这就意味着,在分析数据前,我们就需要做许多文献工作。我们根据需要的指标变量得到一组数据后,立即就可以进行数据分析了吗?许多有经验的分析者的答案是,不可以。我们需要了解这组数据的收集和测量的方法等信息,需要评估数据组的质量。在分析之前,我们可能仍然需要做大量的数据清洁工作。我们的数据分析可能并不需要数据组的全部个案,那么我们就得从中选取适当的个案。有时,我们在分析前就需要处理缺失值的问题。诸如此类的工作并不是一蹴而就的,实际上,也包含了大量取舍上的困难。这些工作既是规范的,也是经验的,要根据实际情况来应变。

第二,我们需要审慎地设定模型。在做一组变量的回归分析之前,我们可能需要花一些时间来处理变量的问题。比如,是否需要对数据中的变量加以变换,生成一个新的变量,作为回归的因变量或自变量呢?是否需要对变量重新编码,改变变量的测量层次,从而得到更有理论意义或更方便解释的结果呢?我们可能还需要通过列联表、直方图、散点图等方法,在分析前就探寻一下用以分析变量间的关系。我们如何确定选取进入实际分析模型的变量是适当的呢?无论是根据理论的需要,还是根据统计的需要,这都是非常棘手的问题,要在以往经验的基础上,审慎地选择自己的策略。选取了实际分析的变量后,我们还要设定变量间的函数形式,比如,如何处理非线性问题,如何处理交互效应等。随机部分的假设也是模型设定的重要组成部分,也需要我们清晰且慎重地设定。

第三,我们需要做好回归诊断工作。回归诊断包括,特异值的发现和处理,有影响的观测值的发现和处理,以及多重共线性的考察和处理。这就包括多种相关统计量的计算、理解和取舍。这些工作也涉及一个好的研究发现和实际分析工作间的平衡,而这又取决于个人的技能和经验。

第四,我们需要合理地解释分析的结果。以回归为例,回归的判定系数越大,我们的模型就拟合得越好吗?是否还存在其他更有效的拟合优度的统计量呢?如何判定某个自变量的效应是否存在呢?如何区别该变量的实质上的显著和统计上的显著呢?我们能够比较两个自变量效应的大小吗?在什么样的条件下,我们才能合理地比较效应的大小?我们可能需要非常审慎地用手头的数据分析做一般的统计推论。我们可能更需要审慎地用回归分析做因果推论。我们需要清晰地理解回归分析的性质,明白它能做什么和不能做什么,以及回归分析同理论验证间的关系。

回归分析像所有其他统计分析工具一样,是一项需要工匠技能的工作,而不是一件可以

随便使用的工具。最重要的是,我们不能随意地把数据送入软件,然后等待结果,再不加思索地声称"结果证明了我的理论"。

4.田野研究

田野研究是另一种有别于抽样调查的综合式调查研究方法。这种方法要求我们在社会情境、行为和事件发生的自然背景、文化背景和制度背景下,研究情境、行为和事件的实际发生和发展情况,从而获得关于某个具体的生活领域或具体的行动领域的理论知识。

田野研究高度依赖研究者自身。它需要研究者具有训练有素的眼睛、耳朵和嘴巴,具有相应学科的理论素养,并以此为工具,到现实的社会环境中,看、听、问和思。只有具备实际田野技能的研究者才可能在田野中获得理论发现。此外,田野研究对研究者而言也不仅仅是发现的场所,它能够赋予研究者其他研究方式难以提供的鲜活体验。这些鲜活体验就像酵母一样,不仅对我们的当前研究,还会对我们的后续研究,产生巨大的作用。

田野研究需要我们实践多种研究方法,掌握多种研究技能。我们在做田野研究时,通常需要掌握档案研究的技能、参与观察的技能、访谈的技能和问卷调查的技能。

简单地说,在田野研究之前,我们可能需要查阅大量的官方文档、组织记录、网络资料等,以帮助我们获得调查地的历史和对当下的初步认识,赋予我们关于调查地的一些地方性知识,以及研究开展的初步线索。在田野研究中,我们可能需要随时收集和查阅一些档案资料,补充由其他方法获得的调查资料。在田野研究后,我们可能需要为调查地建立起自己的档案资料库,既为了以后的研究便利,也可以开放给其他研究者使用。

参与观察是田野研究中一种主要的研究技能。无论作为完全的参与者,还是作为观察者的参与者,研究者都得具备相应的技能。比如,如何进入观察情境,融入观察情境?如何克服视而不见的习惯?如何提高我们观察的敏锐度?如何保持我们观察的注意力?如何增进我们对各种情境、行为和事件过程的清楚记忆?在田野中,如何记笔记,做简记、日志和阶段小结?如何在田野中保持旺盛的精力,克服田野的孤独感,避开田野中的危险等,也一样是参与观察者可能需要掌握的技能。

访谈是田野研究中另一种主要的研究技能。简单地说,田野研究中有两种访谈,深度访谈和焦点小组访谈。两种访谈可能都需要研究者在事前精心准备访谈提纲,无论是无结构的,还是半结构的。深度访谈需要访谈者具有倾听、追问,以及准确理解被访者谈话意图的技能。在焦点组访谈中,访谈者也是主持人。这时,访谈者就需要掌握控制访谈场景和访谈节奏的技能。

田野研究并不排斥问卷调查,就如同抽样调查不能排斥田野研究一样。两者在研究者的调查实践中不仅应当结合起来,而且需要更好地结合起来。因此,从事田野研究的研究者还必须掌握问卷调查和数据分析的相关技能。

最后,田野研究不仅是技能,还是伦理。研究者可能需要谨慎地控制自己在调查地扮演的角色,无论作为法官、检察官、行动者还是研究人员。研究者还可能需要始终如一地记住自己对调查对象所担负的责任和义务。

5.敏化概念

作为社会科学理论的学生,我们的主要困难是缺乏用以理解抽象概念的实际经验和背景知识。作为日常生活中的一员,我们又常常陷入到生活的种种经验细节之中,无法从中获

得相应的理论知识。那么一个研究者如何可能具有对经验世界理论感知，并能够从经验世界走向理论发现呢？美国社会学家布鲁姆提出了“敏化概念”。简单地说，敏化概念就是要求研究者深入经验世界之中，从而赋予抽象概念以感知力，同时用具有感知力的概念来引导研究者对经验世界进行系统探究，从而达到理论精炼或理论发现的目的。

敏化概念的掌握可以从两个方向进行。一个方向是用理论概念审视经验世界，另一个方向是通过经验材料的系统编码来走向理论概念。

用理论概念审视经验世界就是主动把理论概念当作我们自身感知器官的组成部分，去看、去听和去感受。举例来说，官僚人格是研究官僚制会涉及的一个理论概念。研究者可能已经从他人的文献中学会怎样去界定、分类或操作这一概念，却未必会用这个概念来审视他们的周遭世界。这时研究者可能需要暂时放下关于官僚人格的学术定义，而是实际地看一下官僚人格的发生。比如，读一读王蒙的小说《组织部来了个年轻人》。这篇小说的作者透过事件发生的场景、过程以及人物的对话，描绘了多种官僚人格。这些官僚人格不再是理论世界的内容，而是现实生活的组成部分。当然，研究者可能更需要在他自己的日常生活中发现官僚人格，发现官僚人格的各种表现和类型，发现单一的和复杂的官僚人格，发现与官僚人格相冲突的其他人格等。当研究者在他的日常生活或田野中，能够敏锐地察觉到各种官僚人格的展现时，官僚人格就不仅仅是外在于他的理论概念，而是他的上手工具。

研究者也可以试着按扎根理论的要求去做，暂时悬置已有的理论概念，保持理论发现的追求，沉入到经验资料中，通过由低到高的抽象思考，通过资料间的恒定比较，通过编码和写作备忘录，争取“浮现”理论。在编码过程中，研究者先用不同的初级编码来确定经验资料的差异，再通过初级编码的归类走向更高层次的理论概括。编码过程先是揉碎经验资料的过程，再是组成概念图像的过程。研究者可以把编码和写作备忘录看作一种系统的训练，因为这样做的确可以培养研究者的理论敏锐。

6.个案研究

个案研究是另一种经常为研究者使用的研究策略。研究者还可以把它同定量研究、实验研究或模拟研究结合起来使用，用多策略研究方式寻求理论发现和理论检验。那么如何从社会和历史的现实中取出一个个相对独立的，既具有经验的独特性又具有理论的普遍性的经验对象呢？又如何在构建“个案”的同时，精炼相关的理论呢？

首先，研究者可能需要具有界定个案的技能。初看起来，个案的界定似乎不成为问题，但是研究者如何仔细地思考一下“什么是个案”，就会发现这个问题难于回答。特别是在实际研究中，研究者在研究设计之处就需要较为明确地界定个案的经验的和理论的边界。个案可能是一个人、一段生活、一种生活样式；也可能是一个社区、组织或制度；也可能是一个个小的事件或具有历史文化意义的宏大事件；也可能是一个个具体的社会情境或模糊的社会趋势。那么如何确定个案经验上的时间、空间边界，以及可能包含的行动者和事件的边界呢？这恐怕难以有现成的答案。许多时候这只是作为问题需要研究者在自己的研究之前、研究之中和研究之后不断地追问，而且研究者也只能凭借着自己训练有素的理论直觉和实际的研究予以解答。

其次，研究者可能需要具有构建个案的技能。个案不是地理空间中一个清楚的地标。个案是研究者自己根据他们的理论探索，从经验世界中构建出来的认识对象。这个经验对

象可能存在于档案之中,存在于人们的言行之中,存在于人们的记忆和心灵中。要将一个个案从中取出来,就需要研究者具有提出具体研究问题的本领。这些具体的研究问题不是向他人提出的,而是研究者向自己提出的。它的形式是,“我想要知道……是何、如何和为何?”只有通过具体的研究问题,研究者才有可能构建个案的骨架和血肉。同时,研究者的研究问题越具有理论意义,他的个案越可能具有理论价值。

再次,研究者可能需要具有追问事实的技能。这里的事实指的是一个人的讲述是否真实,或一个事件发生的实际过程等。研究者不可能凭借单方面的材料来构建经验对象,可能需要运用三角测量的原则,从多个视角出发,运用多种研究工具,并通过多种研究资料的收集和分析,来构建确实可信和丰富饱满的个案。比如,用几个共同参与者的访谈来相互校验、纠正和补充他们的共同经历。

又次,研究者可能需要具有理论精炼的技能。在从经验资料中构建起个案的同时,研究者也在淬炼他们的分析工具。研究者既可以通过呈现个案的独特性来精炼一个概念,也可以通过多个个案的比较来精炼某一特定理论,或者通过对比不同尺度、不同复杂程度、不同组织形式的个案来精炼某一理论框架。研究者既要沉入到个案的经验细节之中,也要从经验事实上升到抽象的云层,做到“一览众山小”。这可能需要研究者掌握分析式归纳的技能,通过归纳走向一般,再从一般回到具体经验,进而将一个个丰富饱满的经验对象变为具有现实感和说服力的理论图示。研究者可能需要用同一件理论工具雕琢出不同的经验对象,也需要不断完善自己的理论工具箱。

最后,个案研究可能需要具有团队研究的技能。在从事多个案研究、小 N 研究、大尺度的复杂个案研究或多视角的个案研究时,研究者可能就需要组成一个团队,一起收集和分析经验资料,一起署名发表。这时就可能涉及许多团体合作的问题。比如,团队领导者的组织力和协调力。如何建立团队的进取精神?如何形成团队的分工和协作?如何处理团队成员间的冲突?如何处理个别成员的搭便车问题?如何协调研究成果的共享和署名权问题等。在团队研究中,团队领导者和成员可能都需要掌握一定的社交技巧,建立一个既能相互激发又能相互尊重的研究氛围。

通过上述六个例子,我们主要想说明,方法是需要掌握的技能,技能是“我”的方法。为了探究我们生活于其中的经验世界,把纷繁复杂的种种经验现象做成合理的、有现实感和美感的理论图示,我们就必须掌握用以感知和思考的工具。只有当我们能够像能工巧匠运用他们的斧和凿一样,规范、熟练,甚至巧妙地掌握自己的手头工具时,我们才能做出“好的”科学成果。当然,我们也只有通过实际的经验研究,才能真正把方法内化为自己的研究技能。

最后,我想再次强调,本书是学习和应用社会科学研究方法的好向导,希望广大读者能在这个向导的引领下,在学习和应用社会科学研究方法方面取得更大的进步。

赵　锋